Consumer Lending Discrimination in the FinTech Era*

Robert Bartlett† Adair Morse‡ Richard Stanton§ Nancy Wallace¶

November 2, 2017

Abstract

Lending discrimination can stem from loan officer facial biases or algorithmic scoring, especially with big data use in FinTech. Using never-before-linked mortgage data covering loan-level ethnicity, scoring variables, contract terms, and lender identifiers, we implement a treatment-based Oaxaca-Blinder discrimination estimation, based on the unique default risk setting of the GSEs. We find that African-American and Hispanic borrowers have a 2% higher loan rejection rate, especially among low-credit-score applicants. Consistent with facial biases, differences are more pronounced among smaller lenders and independent mortgage companies, not FinTech lenders. Ethnic-minority borrowers pay a 0.2% higher interest rate fairly uniformly across lenders, probably resulting from profit-taking opportunities in weaker competitive environments.

Key words: Mortgage discrimination; algorithmic underwriting

*Urban Economics is gratefully acknowledged. For helpful comments and suggestions, we thank Sanjiv Das and participants at the 2017 Boston JOIM Conference and 2017 Annual Research Conference of the Center for International Securities and Derivatives Markets.

†Berkeley Law, U.C. Berkeley, rbartlett@berkeley.edu.

‡Haas School of Business, U.C. Berkeley and NBER, morse@haas.berkeley.edu.

§Haas School of Business, U.C. Berkeley, stanton@haas.berkeley.edu.

¶Haas School of Business, U.C. Berkeley, wallace@haas.berkeley.edu.

1 Introduction

Credit scoring has long been the first step in lenders' processes of approving and pricing household loans. Recently, technology-enabled "FinTech" loan companies have sought to drive significant cost reductions in the underwriting process by improving scoring precision with big data algorithms. Beyond costs, algorithms reduce the influence of humans, who can produce discriminatory loan decisions with respect to race, gender, and ethnicity. However, algorithms have a potential dark side; by using big data to proxy for legitimate repayment ability variables (e.g., hidden wealth), they may introduce illegitimate statistical discrimination. This potential dark side has not gone unnoticed. The potential for illegitimate statistical discrimination toward protected classes of borrowers was a key aspect of Congressman Emanuel Cleaver's 2017 investigation into FinTech lending. Whether FinTech solves an age-old problem of human prejudice or introduces algorithmic discrimination remains an open question, a question that will only grow in importance as more data becomes readily available.

In this paper, we estimate the level of ethnic discrimination among conventional and FinTech lenders. We focus our examination on the mortgage market, which represents 71% of the $12.73 trillion household loan market as well as the largest market for FinTech loans.[1] The analysis uses a data set that includes never-before-linked information at the loan-level on income, ethnicity, debt-to-income ratios, loan-to-value ratios, all contract terms (such as coupon, loan amount, installment payment structure, amortization, maturity, loan purpose, and mortgage-origination month), and indicators for whether the lender-of-record primarily used algorithmic scoring.

To identify the presence of ethnic discrimination, we first need a landscape of what is legal and what is not. Discrimination comes in three types. Direct illegal discrimination (termed *disparate treatment* in law) happens when lenders' loan-offer decisions differ explicitly by ethnicity, either in their algorithmic processes (i.e., ethnicity is a scoring variable) or in personnel biases (including both cognitively aware and unaware biases). Redlining, the use of geography to approximate for ethnicity, is also deemed directly illegal under *disparate treatment*.

The second and third types of discrimination emerge from statistical discrimination techniques. Imagine a lender attempting to evaluate a loan applicant's repayment risk. Variables which structurally affect repayment in expectation are life cycle model variables – income, debt, wealth, expenses, and derivatives from these variables. These are legitimate (struc-

[1] See, https://www.nytimes.com/2017/05/17/business/dealbook/household-debt-united-states.html.

tural) variables for lender algorithms. The lender cannot, however, observe all of these variables; for example, family wealth is difficult to observe. Thus the lender relies on other observable characteristics to proxy for hidden family wealth. Clearly, if the lender uses ethnicity as a proxy, the lender is directly violating the law under disparate treatment. More likely, however, the lender may instead proxy with other characteristics, such as education or marital status. These variables may be correlated with ethnicity. To the extent that the correlation of education with ethnicity is an artifact of the correlation of education with hidden wealth or hidden income growth, this is legitimate statistical discrimination.

Additionally under existing law, a lender would have the opportunity to justify its disparate impact by demonstrating the strength of its correlation with credit risk (e.g., its correlation to structural life cycle variables). However, if part of the correlation of education with ethnicity is orthogonal to the structural life cycle variables, then the use of education implies some illegitimate statistical discrimination. As big data use increases, the problem gets worse. Algorithms could easily be implemented to predict default based on the exact college or high school one attended. Likewise, whereas the number of children may predict expenses, a convex algorithm on children may predict default that weights heavily on ethnicity, unrelated to direct child expenses, which should be concave.

Under the law, this last use of statistical discrimination—that is, where a scoring variable is largely orthogonal to structural life cycle variables—can create liability for lenders under the legal theory of *disparate impact*. The term conveys the idea that a scoring process disparately impacts an ethnic group, unrelated to structural fundamentals. The challenge posed by FinTech's use of big data is that if a variable predicts default at all, a lender may deem the variable to have credit risk assessment validity. However, one can imagine that the indicator vector of attendance in all possible high schools or colleges could predict default, but would also pick up ethnic sorting unrelated to structural credit risk.

This paper's identification strategy is the Oaxaca-Blinder decomposition method of the labor literature, with the modern implementation in a treatment frame (Fortin, Lemieux, and Firpo, 2011; Kline, 2011). Our outcome variables are the lenders' accept/reject decision and the interest rate borrowers face conditional on acceptance. The method decomposes ethnicity effects that emerge because of differences in the life cycle covariates (legitimate statistical discrimination) versus differences that are either disparate treatment (explicit bias) or disparate impact (illegitimate statistical discrimination). Treatment is being in the African-American or Hispanic ethnic groups.

Our empirical challenge is asserting that any residual differences among ethnic groups is not due to lenders' use of legitimate life cycle variables unobservable to us. Our identification rests on framing our tests within the underwriting and pricing standards that are imposed on the conventional conforming securitized loan market by the Government Sponsored Enterprise's (GSEs). It is well known that, post-crisis, the GSEs (Fannie Mae and Freddie Mac) purchase and securitize more than 90% of the conventional conforming mortgage market in the U.S. It is less recognized that, post-crisis, it is the GSEs who dictate the allowable combinations of contract and eligible borrower characteristics. Once a mortgage clears the credit quality assessment of the GSEs based on observables (e.g., income-to-debt ratios, the loan-to-value ratio, credit scores), lenders face only documentation risk (no residual credit risk) that a future defaulting loan will be "put-back" to the lender. Because the only variables exposed to this risk are observables, our empirical design for accept/reject estimations satisfies the mean conditional independence assumptions of being able to disentangle legal from illegal/illegitimate discrimination.

Our pricing estimations likewise benefit from the GSE structures. The GSEs produce a predetermined grid pricing that prices credit risk across loan-to-value and credit score buckets. Any deviation from this grid pricing must reflects lenders competitive agenda in capturing volume or profit per mortgage. Any ethnic differences in pricing emerging from competitive position would not satisfy the courts as legitimate business necessity. Therefore, herein again, the GSE structure allows us to identify illegitimate discrimination, satisfying omitted variables concerns. For these reasons, we focus our tests on a large dataset of loan applications and approved loans that are identified as being sourced to a GSE between 2007 and 2012.

We find that African-American and Hispanic applicants are 2% more likely to be rejected for a mortgage than other applicants. These higher rejections are most pronounced in low credit score buckets. We interpret this result that lenders have illegal biases against ethnic minorities in terms of assessing the reliability of documents. Consistent with this interpretation, we find that it is small, independent mortgage originators, not large banks and not FinTech firms that impose greater discrimination. This is consistent with loan officer facial bias; these loan officers need not be cognitively aware.

In loan pricing, we find that accepted borrowers in Hispanic and African-American ethnic groups pay a slightly (0.18%) higher interest rate. These results are more uniform across lender types and within credit scores. The differential price result, albeit small, suggest that either lenders try to generate more volume among white majority borrowers by offering rate discounts or that lenders feel the minority borrowers shop around less, either because of personal characteristics or because of financial services deserts, affecting the ability of lenders to impose monopoly pricing.

The paper is organized as follows. In Section 2, we discuss anti-discrimination regulations in U.S. mortgage lending, mapping the law to economic methods in statistical discrimination.

We present our methodology for the measurement of mortgage discrimination in Section 3, and introduce the role of the GSE pricing and scoring grids in practice. In Section 4, we discuss our multifaceted data. Our empirical results are reported in Section 5 and Section 6 concludes.

2 Legality of Explicit and Statistical Discrimination

Discrimination in residential real estate lending is policed primarily by two federal statutes, the Fair Housing Act of 1968 (FHA) and the Equal Credit Opportunity Act of 1974 (ECOA).[2] The FHA, which is administered by the Department of Housing and Urban Affairs (HUD), prohibits taking any action affecting the terms of the transaction on the basis of a borrower/applicant's race, religion, national origin, gender, familial status (i.e., family size or marital status), or handicap. The ECOA builds on the FHA by expanding the set of protected classes of borrowers. Under these laws, a lender in the mortgage market may not engage in, among other things, (i) refusing to extend credit, (ii) using different standards in determining whether to extend credit, or (iii) varying the terms of credit offers (e.g., loan amount, interest rate, duration and type of loan) on the basis of one of the above-mentioned protected characteristic.

Under the FHA and the ECOA, either a wronged borrower/applicant or the relevant administrative agency has the authority to bring a civil action against a lender. As in other areas of anti-discrimination law, however, the scope of these anti-discrimination mandates has been determined largely by the legal standards that courts have created for a successful claim. Historically, the primary method by which lenders were found to have engaged in prohibited discrimination was by a showing of *disparate treatment* — specifically, a showing that a lender had treated borrowers or applicants differently because of a protected characteristic. For example, in Watson v. Pathway Financial, 702 F. Supp. 186 (N.D. Ill. 1988), an African-American couple successfully sued a lender under the FHA for rejecting their mortgage application because of delinquent credit card accounts. While the lender's justification was a potentially legitimate basis for denying credit, the court found the lender had violated the FHA because the applicants produced evidence that the lender had approved at least six applications from white borrowers showing similar delinquencies.

Disparate treatment claims have also been used to prohibit traditional redlining, in which

[2] Although they lack explicit anti-discriminatory prohibitions, the Community Reinvestment Act of 1977 (CRA) and the Home Mortgage Disclosure Act of 1975 (HMDA) also seek to curb discriminatory lending practices by, respectively, mandating a certain level of lending in low- and moderate-income neighborhoods and requiring public disclosure of mortgage data as it relates to ethnicity in order that the public can monitor for discriminatory patterns.

a mortgage lender refuses to make loans to entire geographic areas because of their racial composition (see Gano, 2017). Thus, for economists, the two sets of variables that are explicitly illegal under disparate treatment are indicator variables of the protected category (e.g., ethnicity in our case) and geography.

In addition to disparate treatment claims, a private party or governmental agency can bring a claim of lending discrimination under the FHA or ECOA under a *disparate impact* theory. In contrast to disparate treatment claims, claims of disparate impact do not involve any allegation of intentional discrimination in how a lender treats applicants/borrowers but rely instead on the fact that lending practices that are facially neutral in their treatment of different groups nonetheless fall more harshly on a protected category of applicants. For instance, in a joint policy statement on the enforcement of the ECOA and the FHA, the Department of Justice, HUD, and all federal banking regulators provided the following example as an illustration of a lending policy that could give rise to a disparate impact claim:[3]

> *Example.* A lender's policy is to deny loan applications for single-family residences for less than $60,000. The policy has been in effect for ten years. This minimum loan amount policy is shown to disproportionately exclude potential minority applicants from consideration because of their income levels or the value of the houses in the areas in which they live.

Despite agency approval of disparate impact theory, the ability of parties to pursue disparate impact claims has been hindered for two reasons. First, the existence of a disparate impact, by itself, is insufficient to prove illegal discrimination. Rather, the lender can defend the practice as justified by a legitimate business necessity, provided no alternative policy or practice could achieve the same goal with less discriminatory effect.

Second, the statutes fail to specify a standard for proving a discriminatory effects violation. What the standard of proof should be is not obvious, and regulators simply lack processes. Under FHA and ECOA, enforcement agencies make clear to regulated lenders that they are required to evaluate policies and procedures for evidence of disparate impact, but little is done in practice. In fact, it was not until the Supreme Court's 2015 decision in *Texas Department of Housing and Community Affairs v. Inclusive Communities Project* that the Court even approved the disparate impact framework under the FHA. Yet even after the Court's ruling, enforcement agencies have yet to revise their prior enforcement guidance with respect to fair lending examinations.

The economic literature has something to contribute to the question of specifying a stan-

[3] See Policy Statement on Discrimination in Lending, 59 Fed. Reg. 18,266 (Apr. 15, 1994).

dard of proof.[4] Starting in the 1970s (see Aigner and Cain, 1977; Arrow, 1973; Phelps, 1972), economic research shifted discrimination discussions to the statistical theory of discrimination rather than taste-based discrimination associated with Gary Becker. The language of statistical discrimination maps well to a legal standing under disparate impact. However, economists have not been able to substantiate the link, because of a lack of progress in positing causality in statistical discrimination estimation. Criticisms of inadequate data and omitted variable biases in estimation have plagued this literature.[5] For example, Shafer and Ladd (1981) find some evidence of interracial pricing differences, and Black and Schweitzer (1977) find indications of differences in loan terms. Yet Sandler and Biran (1995) critique any use of statistics in legal proceedings of discrimination claims due to the prevalence of omitted variable bias and poor identification strategies in these studies. Not surprisingly, the mortgage literature on discrimination parallels these broader patterns, primarily focusing either on the methodological difficulties in providing robust statistical evidence for discrimination (see Black, Schweitzer, and Mandell, 1978; Kaye, 1982; Maddala and Trost, 1982; Rachlis and Yezer, 1993) or on the efficacy of specific types of legislation.[6]

However, we think that more can be accomplished. Consistent with the Supreme Court's decision in *Inclusive Communities*, the theory behind statistical discrimination provides guidance as to how one can demonstrate a "robust" causal connection between lending practices and racial or ethnic disparities in lending outcomes. In statistical discrimination, agents (in our case lenders seeking to screen applicants) have limited information and no animus against racial groups. Statistical discrimination arises as a solution to a signal extraction problem. Agents seek to reconstruct hidden information as to the expected creditworthiness of applicants using observable proxies.

We have already established that the use of the protected variable (ethnicity) and geography are directly illegal under disparate treatment. However, could other variables be legal proxies for hidden information, even if they are correlated with ethnicity? Under disparate impact rulings, statistical discrimination is allowable for "legitimate business necessity." Economic theory guides us that the meaning of this phrase is that a variable legitimately appears as a structural variable that maps the ability of a household to repay a loan to their

[4] See commentary by Glassman and Verna (2016) and Winfield v. City of N.Y., No. 15CV5236-LTS-DCF, 2016 WL 6208564, at *1 (S.D.N.Y. Oct. 24, 2016); Cty. of Cook v. HSBC N. Am. Holdings Inc., 136 F. Supp.3d 952, 955 (N.D. Ill. 2015).

[5] Cited data omissions in empirical tests of discrimination include omitted information such as the loan to value ratio at origination, the debt to income ratio, all the contracting elements of the mortgage, the property characteristics and exact address, the applicant's ethnicity, gender, credit history and debt burden levels, lender and regulatory characteristics, all at the loan level for successful and unsuccessful applicants.

[6] For example, the Fair Housing Act of 1968, the Equal Credit Opportunity Act of 1974, the Home Mortgage Disclosure Act of 1975, the Community Reinvestment Act of 1977, and the HMDA amendments to the Financial Institutions Reform Reregulation and Enforcement Act of 1989.

economic fundamentals. In particular, one can write down a life cycle model in which cash flow for repayments emerge from the current borrowing position (debt), cost of borrowing (credit score), income (in levels, growth, and risk), wealth, and regular expense levels (cost of living measures). Thus, the use of any of these variables should be considered "legitimate business necessity" under disparate impact theory, even if this variable statistically loads on (punishes) a particular protected category. This is legitimate statistical discrimination.

What if a lender cannot see one of these variables, say, wealth? However this lender can see a variable (the name of the high school attended) that correlates with wealth. Under disparate impact, this variable should be allowable if it is only disparately impacting the pool of applicants in sorting on wealth. In other words, conditional on latent hidden wealth, high school is orthogonal to loan decisions. A slightly less stringent assumption would also be consistent: conditional on latent hidden wealth, the impact of high school on loan decisions must be the same for one ethnic group as the other (ignorability).

3 Methodology

3.1 Overview

Modern methods to estimate discrimination emerge from the works of Oaxaca (1973) and Blinder (1973). Oaxaca and Blinder established a decomposition method whereby ethnic (or gender) differences in an outcome variable are decomposed into that which can be explained by differences in structural covariates and that which remains unexplained. The first piece is selection; the latter is discrimination. In our context, the outcomes of interest are, iteratively, whether the loan accept/reject decision and loan pricing conditional on acceptance. Structural covariates, (i.e., legitimate sources of variation for lenders to use in scoring) are life cycle variables that are truly structural sources of repayment risk assessment.

Decomposing selection from discrimination is not as straightforward as simply applying the Oaxaca-Blinder decomposition because, as we discuss in this section, conditional independence is hard to assert. We exploit the institutional framework by which GSEs underwrite loans and price risk to allow us to satisfy the independence conditions.

3.2 Oaxaca-Blinder in an Ethnicity Treatment Frame

Recent innovations cast the Oaxaca-Blinder methodology in the treatment effect language of the program evaluation literature (Fortin et al., 2011; Kline, 2011; Słoczyński, 2015). We follow in this tradition. The treatment group in our study is African-American and Hispanic mortgage applicants. Our control group includes all other ethnicities.

Our objective is to estimate the average treatment effect on the treated (ATT); i.e., the difference between the realized loan outcomes for ethnic minority mortgage applicants and the counterfactual outcomes for the same applicants if they were credit scored in the pool of the control ethnicities (predominantly Caucasians). We refer to this counterfactual using the terminology "under blinded credit scoring" because our control group represents a large majority of mortgage applications. Our first assumption is just that — that we do not suffer from the lack of a base group, a criticism of the Oaxaca-Blinder methods (Oaxaca and Ransom, 1999); namely:

> *Assumption 1 (Simple Counterfactual for Treatment): Credit scoring of an ethnic minority household would be the same as scoring of the control group scoring (blinded scoring) if households were not identified to be in the ethnic treatment group.*

Next we set up the decomposition. Let subscript T stand for the treated (African-American and Hispanics) and B for the control (B denoting blind). Each household i has a potential loan outcome associated with being credit scored as if an ethnic minority Y_{Ti} and being credit scored Y_{Bi} in a blind model. Only one of Y_{Ti} or Y_{Bi} is ever observed for any household i. We are interested in the difference in the average values

$$\Delta = \overline{Y}_T - \overline{Y}_B$$

Next, we make the standard linearity assumption in this literature:

> *Assumption 2 (Linearity in Structural Variables): Outcomes Y_{Ti} and Y_{Bi} are linearly related to the structural variables, denoted X where X is the vector $(X_i = [X_{i1}, ..., X_{iK}])$.* [7]*:*

$$Y_{Ti} = \beta_{T0} + \sum_{k=1}^{K} X_{iTk}\beta_{Tk} + \varepsilon_{Ti}$$
$$Y_{Bi} = \beta_{B0} + \sum_{k=1}^{K} X_{iBk}\beta_{Bk} + \varepsilon_{Bi}.$$

[7]Lenders construct credit scoring models from discriminant analysis of defaults. We anecdotally assume that these models linearly analyze structurally-implied variables or step functions of these variables. The scoring models emerge from such analyses, allowing us to approximate with linear terms. We do nonparametric versions of our estimations to show robustness to this linearity assumption.

The difference Δ in outcomes can be decomposed as follows:

$$\Delta = \underbrace{\sum_{k=1}^{K}\left(\overline{X}_{Tk}-\overline{X}_{Bk}\right)\widehat{\beta}_{Bk}}_{explained} + \underbrace{\left(\widehat{\beta}_{T0}-\widehat{\beta}_{B0}\right)+\sum_{k=1}^{K}\overline{X}_{Tk}\left(\widehat{\beta}_{Tk}-\widehat{\beta}_{Bk}\right)}_{unexplained}.$$

Loan outcome differences result from (i) (explained) the difference in the household structural variables, applying the blinded credit scoring, (ii) the difference in the constant shifters in outcomes, and (iii) the difference in outcomes due to the treated being scored with different scoring parameters, relating the structural variables to loan outcomes, compared to the blinded scoring parameters. The sum of the two unexplained terms is the total discrimination.

We need two additional assumptions to posit that the decomposition can causally identify discrimination. These assumptions are at the heart of the role GSE grids play in our methodology.

> *Assumption 3 (Overlapping Support): Each possible treated realization of* $X_i = x$ *and* $\varepsilon_i = e$ *must be in the common support; i.e.,* $0 < \Pr\left[i \subseteq T \mid X_i = x,\ \varepsilon_i = e\right] < 1$.

> *Assumption 4a (Conditional Independence):* Applicants' unobserved life cycle characteristics are independent, conditional on observed covariates. $E\left(\varepsilon \mid X\right) = 0$

Kline (2011) establishes that a lighter version of Assumption 4a is all that is necessary in this setting:

> *Assumption 4b (Ignorability): Any selection based on unobservables must be the same for the treated and control. Unobservables do not need to be independent of X, as long as their distribution conditional on X is the same for both ethnic groups. Denoting the distribution function by* $g\left(\cdot\right)$, $g\left(\varepsilon \mid X, i \subseteq T\right) = g\left(\varepsilon \mid X, i \subseteq B\right)$.

Neither 4a nor 4b is easy to establish in the setting of applicants selecting to apply for a mortgage. The most worrying concern is the implications from omitted life cycle variables. Imagine that expected income growth causally relates to repayment risk. However, expected income growth is unobservable. Furthermore, it is correlated with current income, but the intensity of this correlation is different for ethnic minorities as compared to the ethnic majority. All of that seems very plausible. In this case, the omitted variable bias in the scoring model parameters would lead to a biased Oaxaca-Blinder decomposition. Moreover, the weaker ignorability assumption does not help because the βs would not be biased in the same magnitude.[8] Fortunately, these concerns are greatly attenuated in the context of GSE lending.

3.3 Discrimination in Loan Rejections & Error Independence

The GSEs play an important MBS market role of buying mortgages from originators, packaging mortgages into MBS, and guaranteeing loan payments to MBS investors. We argue that the guarantee of loan payment provides us with exogeneity to unobservables that allows for identification of discrimination. While an originator who retains a mortgage in its portfolio is incentivized to estimate an applicant's credit risk, an originator who is engaged in sourcing and selling loans to the GSEs ordinarily bears no credit risk, except for *put-back risk* discussed below. Moreover, the criteria used by GSEs for determining whether a loan is acceptable for purchase and pooling are based entirely on observables characteristics.[9] Within a set loan type (i.e., 30-year fixed contract for an owner-occupied, single family home), Fannie Mae's Sellers' Guide specifies these characteristics to include: Mortgage Eligibility (maximum LTV, allowable loan amortization), Borrower Eligibility (documentation, LTV maximums, credit score minimums), and Occupancy/Property Eligibility (documentation of status).

In addition, the applicant's characteristics are processed for acceptability in an automated underwriter system (for Fannie Mae, *Desktop Underwriter*). The underwriter system considers only the following: an applicant's credit report, liquid reserves, total expense ratio, and co-borrowers. For lenders not using the underwriter system, the Sellers' Guide also specifies that manual underwriting should focus on these same observables. Thus, for our purposes, acceptability of an application is based on observables, implying that there would be no point in a lender engaging in statistical discrimination for default risk.

[8] Note that this omitted variable bias may be even larger for studies which proceed with more descriptive models of discrimination.

[9] The 2010 Sellers' Guide for Fannie Mae, states the following condition for the purchase of loans:

> **Loan Qualification**. Lender must ensure that all loans selected for delivery meet Fannie Mae's underwriting and eligibility guidelines and legal requirements and match the terms of the commitment or for MBS loans, the pool purchase contract, including mortgage type, amortization, original term, and pass through rate(s) selected for delivery when the commitment or contract was created.

Our setting involves one further complication. The lender remains subject to the residual risk that it may have to repurchase a defaulting loan under its representations and warranty liability (*put-back risk).* The importance of this representation was underscored following the Financial Crisis in 2010 when GSEs put back \$4.2 billion of loans (*American Banker*, July 14, 2016).[10] However, the formulaic character of GSE underwriting ensures that both the original accept/reject decision and any subsequent liability for a breach of a representation/warranty will hinge on *observables.* Put-back risk only concerns risks in the documentation and verification of observables (income, debt-to-income, credit score, etc.). Thus, in the current environment of verification and documentation, *any* discrimination based on reliability is unwarranted explicit ethnic bias with no economic basis.

3.4 Discrimination in Loan Pricing & Error Independence

In addition to establishing underwriting standards, the GSEs also play a very significant role in the pricing of mortgages. Each GSE guarantees the timely payment of principal and interest on its MBS and charges a fee for providing that guarantee. These *guarantee fees* (or G-fees) cover projected credit losses from borrower defaults over the life of the loans, administrative costs, and a return on capital to the GSE. A standard G-fee is assessed on all mortgages as a percentage of the loan balance and is collected monthly (see Fuster, Goodman, Lucca, Madar, Molloy, and Willen, 2013).[11]

In March 2008, the GSEs introduced up front fees based on an 8 x 9 matrix of LTV and credit score buckets, called the *Loan Level Price Adjustments* (LLPAs).[12] Figure 1 depicts a typical LLPA grid. These fees have been adjusted four times between 2008 and 2012 in response to changes in the GSEs' forecasts of house price dynamics and credit losses. In practice these one-time fees are commonly converted into monthly "flow" payments that are added into interest rate as rate pass-throughs to borrowers.

Lenders know the grid at any point in time and thus price consumer-facing interest rates that are sufficient to cover these fees as well as any profits (Fuster et al., 2013). Figure 2

[10] A large number of put-backs by a lender will induce the GSEs and the MBS market to add a discount to the price for products from this lender. These overall lender adjustments can be controlled for in our Oaxaca-Blinder methods with firm fixed effects, as in Abowd, Kramarz, and Margolis (1999) and Card, Cardoso, and Kline (2016).

[11] The actuarially fair pricing of the G-fees is also a central policy question in the determination of the future role of the GSEs in the U.S. mortgage markets (see Elenev, Landvoigt, and van Nieuwerburgh, 2016; Vickery and Wright, 2013).

[12] In March 2008, the GSEs also introduced an up-front adverse market charge of twenty five basis points for all loans in order to protect against the heightened credit risk posed by rapidly deteriorating housing market conditions. The LLPA schedules are published by the GSEs and reviewed by the Federal Housing Finance Agency on an annual basis (see Federal Housing Finance Administration, 2009, 2010, 2011, 2012, 2013; Fuster and Willen, 2010).

represents a schematic for how a lender would think about the pricing of a loan contract that is being originated with the intent to securitize through a GSE. The figure represents two mortgage applications with the same credit score and the same loan-to-value (LTV) ratio, both of which have been approved by the underwriting system. The mortgage interest rate that a household sees on a mortgage consists of three parts. First, all mortgages within the same LTV/credit score bucket face the same market price, determined by the Base Mortgage Rate, which reflects the primary market interest rate for loans to be securitized by the GSEs. Second, all mortgages within the same credit score/LTV bucket face the same LLPA fee added onto the market price.

The third part of the pricing is discretionary to the originator and is determined by the lender's ability to originate loans at interest rates in excess of the minimum required to cover the latter two components. Lenders determine offered mortgage rates using internally produced *rate sheets*, which are schedules of possible rates that a loan officer (or algorithm programmer) can charge for individuals within a cell of the LLPA grid. They use discretionary pricing in rate sheets to incorporate strategic pricing for dynamic industrial organization considerations; i.e., monopoly rents and volume positioning in markets. In addition, rate sheets may be simply used to discriminate against either certain types of borrowers who are known to shop around less or certain borrower areas where finance deserts or collusion make sustained higher prices possible.

Imagine some of these differential pricings to the consumer are correlated with ethnicity. (They may all be.) Again, the standard of proof for a disparate impact claim is that a claimant must prove (i) that discrimination occurred and (ii) that no legitimate business necessity mandated the use of a sorting mechanism that discriminated. We translate these requirements into economics in two parts. First, we are able to identify statistical discrimination in interest rates charged to borrowers via the decomposition method. Second, there is no omitted variable, unobservable, which could be part of a structural model that is a legitimate business necessity. In the accept/reject decisions, credit risk is a legitimate business necessity. However, in rates, profit margins cannot claim that status. Thus, all statistical ethnic discrimination which we are able to isolate within the grid represents illegitimate statistical discrimination or explicit bias.

4 Data

4.1 Data Sets

A key obstacle in prior empirical mortgage discrimination studies has been the reliance on the Home Mortgage Disclosure Data (HMDA),[13] which is the only data source with loan-level information on applicant ethnicity for both successful and unsuccessful loan applications. HMDA also contains applicant income and (nonstandardized) information on lender name. What HMDA lacks is information on the contracting structure of the loan (date, interest rate, maturity, loan-to-value ratio) on the type of loan (fixed, ARM, purchase-versus-refinance), on the property characteristics (location, owner occupied, residential-versus-multifamily, etc.), and on the applicant (credit score, debt-to-income ratio, etc.). A challenge with mortgage loan data in the U.S. has been the lack of a linking of HMDA data with other datasets that do contain loan contracting elements, lender and borrower characteristics, and property characteristics.

Thus, we embarked on a multi-year project of linking loan-level data across four separate data sets using machine learning techniques. The data sets in this large-scale statistical merge are:

- HMDA (30.6 million loans and 11.6 million loan rejections between 2008 and 2012). HMDA data include information on borrower income and ethnicity and lender identifiers for loan-level mortgage origination data. Geography is available only at the level of the census tract.
- Dataquick (30 million loans between 2008 through 2012). Dataquick data provide transaction and assessor information including all mortgage lien recording data, loan performance data (i.e. prepayment and default), lender and borrower names and addresses but very little information on actual mortgage contract terms other than the loan amount.
- McDash (28.7 million loans between 2008 and 2012). McDash data provide loan-level data compiled by Black Night Financial Services and includes quite comprehensive information on the mortgage contract elements and loan types.
- Equifax (20 million loans between 2008 and 2012). Equifax data provide (for a subset of loans in the McDash data) information on other consumer financing balances that are held by borrowers in addition to their mortgages (this merge was done by Equifax).

[13] The HMDA surveys account for approximately 90% of mortgage origination in the U.S. (see Engel and McCoy, 2011). HMDA reporting is not required for institutions with assets (when combined with the assets of any parent corporation) that are below $10 million on the preceding December 31, or institutions that originate 100 or more home purchase loans (including refinancings of home purchase loans) in the preceding calendar year (see `http://www.ffiec.gov/hmda/pdf/2010guide.pdf`).

We describe our merging algorithm in detail in Appendix A. To standardize our analysis, we filter the data to focus on 30 year fixed rate, single family residential loans over the period 2008 through 2012. These loans will have all been securitized by the GSEs. We also filter out CRA zip codes. Finally, because our use of median proxies for LTV and credit score, we trim the support to be LTV between 0.3 and 1 and credit scores to be 630 to 770.

For accepted loans, the merge brings a host of property and applicant characteristics from Dataquick, McDash, and Equifax to the baseline HMDA data, which has ethnicity. However, none of these auxiliary datasets covers loan application rejections. Thus, as explained in more detail in Appendix A, we augment the rejection data with proxies for two key scoring variables that are not included in HMDA. We proxy for the applicant's credit score using the median credit score of the census tract location for all Government National Mortgage Association (GNMA) securitzed loans that were originated each year as reported in the McDash data set. A census tract is on average 1600 households (4000 inhabitants), designed by the Census Bureau to reflect relatively uniform economic situation. Likewise, although we have the actual applied-for loan amount, we need to proxy for value of the property to construct a proxy for the loan-to-value ratio. We start with annual snapshots of the property-specific market assessed value and the date of the assessment that are reported in Dataquick for nearlly all residential properties in the U.S. Within each census tract, we take the simple average of these market assessed property values each year. We then use the actual loan application amount as the numerator to construct a proxy for the loan-to-value ratio for a specific property located in that census tract.

4.2 Summary Statistics

In the summary statistics reported in Table 1, FinTech lenders are shown to originate approximately 2.5% of both accepted and rejected loans of GSE 30 year fixed rate mortgages. These data were obtained by matching HMDA lender names to the firms that were identified as FinTech in Buchak, Matvos, Piskorski, and Seru (2017). In contrast, 46.6% of the accepted loans were originated by the top 25 lenders by origination volume in their respective loan origination year and 63.8% of the rejected loans were originated by top 25 mortgage originators. The origination volume rankings were determined by matching HMDA lender names with mortgage origination statistics obtained from *Inside Mortgage Finance*. HMDA reports the ethnicity of the borrower, or borrower application. The table shows that 15.9% of the accepted loans were to African-American or Hispanic borrowers whereas 14.2% of the rejected loans were to African-American or Hispanic applicants.[14] Thus, in raw statistics,

[14]HMDA reports three ethnicity categories: White/Asian, African-American/Hispanic or unknown. For those loans in the unknown category, we applied a classifier for ethnic first and last names using the borrower

our treatment ethnicities have lower rejection rates.

When we run the interest rate estimations, we also want to ensure robustness to debt levels using the Equifax match. Panel C of Table 1 reports the summary statistics for the subsample of 1,121,395 (71% of the total sample of accepted loans), As shown, these loans are quite similar to the loan not matched to Equifax, except that there is a slightly higher share of refinance loans than purchase loans.

4.3 Effect of Grid-Leveling on Outcomes

Before turning to analysis, we present a series of figures showing the effect of leveling applicants/borrowers by analyzing within the LLPA grid. In Figure 3, panels (a), (b), and (c) display histograms of borrower interest rates by treatment (treat =1 is set of Hispanics and African-Americans). Panel (a) shows that in the raw data, both the treated and the control borrowers seem to have bimodal distributions, with a lot of realized interest rates at 5% and then between 5.75% to 6.75%. However, the ethnic treatment group has more mass in the higher of these modes, resulting in higher means. Therefore, in contrast to the summary statistics that Hispanics and African-Americans have lower rejection rates, the histograms reveal that this treatment group fares worse on interest rates in the raw data.

However, when we level the interest rates within the grid, by subtracting out the year-grid cell mean, Panels (b) and (c) show that the differences in rates become much smaller. The ethnic treatment histograms remain, however, shifted to the right (higher rates). Reassuringly, for our robustness checks, Panels (b) and (c) are very similar. Figures 7 and 8 in Appendix B repeat this exercise, showing differences in the same set of histograms by raw rates versus de-meaned grid rates for splits of FinTech versus traditional lenders and then the Top 25 lenders versus smaller brokers.

Figure 4 presents conventional loan data from the Home Mortgage Disclosure Act surveys for loan applications that were rejected. The figure compares loan application rejection probabilities for the raw data (on the left-hand side of the graph) and demeaned by year-LLPA buckets (on the right-hand side of the graph) for borrowers in the control group (ethnicity $\neq$ African-American/Hispanic) and for the borrowers in the treatment group (ethnicity = African-American/Hispanic). In contrast to the rates histograms, the LLPA pricing schedule is helpful in understanding the rejection rates of loans, since the demeaning leads rejection probabilities for treated and untreated loans to be more similar than they appear in the raw data. Similarly, Figure 9 panels (a) and (b) in Appendix C plot the rejection rates over time for FinTech or Traditional (panel (a)) and Top 25 volume lenders or smaller lenders (panel

name on the loan in DataQuick, as applied in Kerr (2008); Kerr and Lincoln (2010).

(b)). FinTech lenders have similar rejection rates as traditional lenders in the grid; however, smaller volume lenders reject far fewer applicants than Top 25 lenders.

In sum, the histogram depictions of the data reveal that within-grid pricing seems relatively uniform across types of lenders and borrowers. However, rejection rates vary substantially, even in the within grid framing. This is perhaps not surprising, as the intent of the grid is to level pricing. Before making inferences, however, we need to implement the full Oaxaca-Blinder decomposition in a multivariate setting.

5 Results

5.1 Rejection Rates

As discussed above, Figure 4 reveals no immediate sign of discriminatory rejection rates, since there is no consistent relationship between either raw or de-meaned rejection rates for the control and treated groups over time. Rejection rates are higher for the control group in some years, and for the treatment group in others. Some patterns do emerge in the data, however, when we split the data in different ways. Because borrower or lender characteristics might differ significantly between the treated and untreated groups, which would make the unconditional comparisons misleading, we now look at rejection rates in more detail.

Focusing first on the full sample of loans for which we have complete data. Table 2 compares regressions of lenders' accept and reject origination decisions on borrowers loan-to-value ratios, their credit scores, and the log of their annual incomes. In column (1), we show the ordinary least squares (OLS) regression results and in column (2) we show the same OLS regression with the inclusion of month and bucket fixed effects. The results for both of the OLS specifications appear to indicate that the effect of the intercept is negative, suggesting that the legal life cycle screening components are closely associated with the rejection of borrower applications and that on average non-legal screening elements, possibly associated with protected characteristics, actually reduce the likelihood of a borrower's loan application being rejected.

In contrast to the OLS results, columns (3)–(5) show results of the Oaxaca (1973) decomposition. As shown, the Oaxaca results suggest that the treatment effect is positive 1.99%. This result implies that, after controlling for observable differences, African-American and Hispanic borrowers are about 2% more likely to be rejected for a loan than other borrowers. As discussed in Section 3 above, the formulaic character of GSE underwriting means that two applicants with similar observable structural variables, and who would be assigned to the same LLPA grid, should face the same likelihood of loan approval. This non-zero treatment effect is clear evidence of discrimination against African-American and Hispanic borrowers. In addition, as shown in column (5), there appear to be statistically significant differences in the weights that lenders attach to the evaluation of legitimate life cycle covariates. These results suggest that lenders overweight the importance of the loan-to-value ratios and credit scores for African-American and Hispanic borrowers and they underweight the importance of the log of their annual income. These differential weights represent another component of illegitimate screening, or discrimination, since the legitimate use of life cycle characteristics should receive equal weighting on average between the two borrower groups.

In the OLS regressions and the Oaxaca decomposition presented in panel B of Table 2 we add lender fixed effects to the panel A specification. As shown, the treatment effect falls only very slightly to 1.79% supporting the prior result that African-American and Hispanic borrowers are about 2% more likely to be rejected for a purchase mortgage through a discriminatory channel. The addition of the lender fixed effects also leaves in tact the results on the discriminatory differential weighting of life cycle characteristics for ethnic minority borrowers where again lenders appear to overweight loan-to-value and credit scores and under weight the log of income.

Tables 3 and 4 show corresponding results for traditional versus FinTech lenders and for top-25 versus smaller lenders, respectively. It can be seen that the discrimination rate for traditional lenders is about the same as for the overall sample (2.08%), whereas FinTech lenders are 7% *less* likely to reject an African-American or Hispanic loan applicant. Here again, as shown in column (3), traditional lenders appear to overweight the loan-to-value ratios and the credit scores of African-American and Hispanic borrowers in their evaluation of these legitimate life cycle characteristics and similar to the overall sample, traditional lenders appear to underweight borrower income. Interestingly, as shown in column (6), there is no evidence of a statistical difference in the weighting of the borrowers' loan-to-value ratios or their credit scores in the FinTech lenders credit rejection decision. There is, however, a statistically significant underweighting of the borrower income for African-American and Hispanic borrowers.

As reported in Table 4 in column (1) and (4) respectively, both large and small lenders show some discrimination, but the differential rejection rate for purchase mortgages is over five times as high for smaller lenders (4.71% versus .8%). The results reported in column (3) indicate that the top 25 lenders in terms of origination volume appear to equally weight the importance of the loan-to-value ratio for African-American and Hispanic borrowers, however, there is a statistically significant overweighting of credit scores and an underweighting of the log of borrower income. In contrast, smaller lenders have a statistically significant overweighting on the importance of loan-to-value for African-American and Hispanic borrowers and a large and statistically significant overweighting on the importance of these borrowers' credit scores. Smaller lenders also have a large and statistically significant underweighting of the log of annual income for African-American and Hispanic borrowers. These results on the presence of large differentials in the weights that small lenders attach to legitimate life cycle characteristics is another concerning indication that smaller lenders are more likely to discriminate against African-American and Hispanic borrowers leading to disparate impact on mortgage market access for these borrowers.

The regressions in reported in Tables 2 through 4 allow different FICO/LTV buckets to have different constants, but otherwise all parameters are assumed to be the same across buckets. To investigate differences across the buckets in more detail, we also run 72 separate Oaxaca (1973) regressions, one per FICO/LTV bucket, and then plot the discrimination coefficient results in Figure 5. Each group of bars representing one FICO bucket, with each bar within one group corresponding to a different LTV bucket. It can clearly be seen here that there is very significant discrimination in rejection rates for low-FICO borrowers, regardless of LTV. Especially for the FICO ranges 640–660, after controlling for observable characteristics, African-American and Hispanic borrowers are between 20% and 30% more likely to have their loans rejected than other borrowers, regardless of LTV. As FICO increases, the extent of this discrimination decreases, to just a few percent for FICO scores between 700 and 720, and turning negative for FICO scores above 720. Indeed, for FICO scores over 740, African-American and Hispanic borrowers are 3%–19% *less* likely to be rejected for a purchase loan than other borrowers.

Overall, these results suggest that African-American and Hispanic borrowers do not have equal access to the purchase mortgage market and that this differential access is due to illegitimate discrimination leading to disparate impact especially amongst smaller lenders who are more likely to rely on face-to-face underwriting. Interestingly, the algorithmic underwriting practices of the FinTech lenders indicate significantly lower incidence of discriminatory credit scoring and a greater reliance on similarly weighted and legitimate life cycle screening factors in their origination decisions for African-American and Hispanic borrowers. For the largest lenders, the top 25 in terms of underwriting volume, we find a treatment effect that is positive and statistically significant but of very small magnitude .8%. This result implies that, after controlling for observable differences, African-American and Hispanic borrowers are have a less than 1% chance of being rejected for a loan compared to the control group of borrowers and this result is in stark contrast with the 4.7% chance of being rejected by smaller lenders and the -7.3% chance of being rejected by FinTech lenders. These results indicate an increased fairness in purchase mortgage lending access from algorithmic underwriting by FinTech lenders and potentially from concerns about exposure to put-back risks on the part of large lenders.

5.2 Mortgage rates

Table 5 shows the results of Oaxaca-Blinder regressions of coupon rates on newly issued 30-year fixed-rate mortgages against various right-hand-side variables. Table 5, panel A, includes the full sample of loans, while panel B reports results for the (slightly smaller) sample of loans merged with Equifax data, allowing us to include additional consumer debt variables (the log of the sum of outstanding monthly balances at origination of student loans, car loans, other consumer debt, and HELOCs) on the right-hand side. In each of the panels A and B, Column (1) shows OLS regressions of the purchase mortgage coupon rate on loan-to-value ratios, credit scores, and log annual income, whereas column (2) includes month and bucket fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of purchase mortgage coupon rate on the same explanatory variables.

As shown in column (3) of panels A and B in Table 5, the treatment effect is approximately .179%. The interpretation of this results is that conditional on being given a loan, African-American and Hispanic borrowers pay an average of .18% more than other borrowers for their purchase mortgage. While this result is not zero and it is statistically significant, it is of much smaller magnitude than the differences in rejection rates identified above. In column (5) of panels A and B, we reported the differential weighting of the life cycle covariates in setting the mortgage contract interest rate. As shown, there appears to be a small, but statistically significant, overweighting of the loan-to-value ratio and an underweighting of the log of the borrower's income. Interesting, lenders do not differentially weight other credit variables in their underwriting decisions.

Table 6 shows results for traditional vs. FinTech lenders. Columns (1)–(3) present the results of Oaxaca (1973) regressions of mortgage rate against life cycle characteristics including month and bucket fixed effects for traditional lenders. Columns (4)–(6) present the same Oaxaca (1973) regressions for the FinTech lenders. As shown in column (3), traditional lenders have a .17% higher coupon rate due to discrimination, which is similar to the value found for the full sample, whereas the discrimination factor is not statistically different from zero for the FinTech lenders. Column (3) reports the differential weighting of life cycle factors among the traditional lenders. Here again there is an over weighting of the loan-to-value ratio and an underweighting of log borrower income but the magnitudes of these differentials is very small. For the FinTech lenders there is a statistically significant underweighting of the log borrower annual income and a marginally significant over-weighting of the log of other debt. Again, the differential magnitudes are very small.

Table 7 shows corresponding results for large top 25 lenders versus smaller lenders. It can be seen that the treatment effect is almost the same for both large and small firms, just under about .2%. The statistically significant differential weightings are also similar and very small.

As previously discussed for the Oaxaca decomposition of the loan origination decisions (the accept/reject decisions), the regressions in Table 5 allow different FICO/LTV buckets to have different constants, but otherwise all parameters are assumed to be the same across buckets. As a further robustness check on our rate regressions, we estimate 72 separate Oaxaca (1973) regressions, one per FICO/LTV bucket, and then compute the mean of these coefficient values by FICO buckets. The FICO bucket means of the discrimination coefficients are reported in Figure 7. Clearly there is very significant trend in the illegitimate discriminatory augmentations to the refi mortgage interest rates as a function of FICO score buckets. The magnitude of the discrimination coefficient decreases monotonically beginning with the 620-640 FICO range indicating that African-American and Hispanic borrowers do pay higher interest rates than other borrowers, regardless of LTV, although the magnitude of the difference is quite small. As FICO increases, the extent of this discrimination decreases, to very small magnitudes for FICO scores greater that 740.

For a final robustness check, we re-run the purchase mortgage regressions that are reported in Tables 5 – 7 using a sample of refinanced mortgages. The results of the regressions for the refinance mortgage interest rates are report in Appendix D, Tables 8 – 10. Overall the results for the Oaxaca decomposition for the mortgage interest rate suggest that the LLPA grids significantly dampens opportunities to discriminate on the basis of the mortgage interest rate. As shown in Panel A of Table 8, the discrimination factor for refi mortgage rates is .29% which is about eleven basis points higher than for purchase mortgages. Similarly, as shown in Panel B of Table 8 this result is unaffected by the inclusion of the log debt. Table 9 reports the results of the Oaxaca decomposition for the regression of refi mortgage rates on legitimate life cycle characteristics for traditional and FinTech lenders. Similar to the results for purchase mortgages we find that the level of illegitimate discrimination is twice high for tradition lenders than it is for FinTech lenders. Interestingly, as shown in columns (2) and (6) of Table 9, traditional and FinTech lenders do not appear to use differential weights on life cycle factors, although the underweight to log income on the part of FinTech lenders is marginally significant. Our final test for evidence of discrimination in the setting of mortgage interest rates for the refi mortgage sample again shows smaller lenders have higher levels of discrimination than the larger lenders, however, the differential magnitudes of this discrimination in rates are less than half 4.7 basis points. Interestingly, in these regressions both traditional and smaller lenders tend to overweight other consumer debt.

Overall, our results for interest rate discrimination indicate that the discipline of the LLPA's leaves very little scope for interest rate discrimination among lenders. On average as shown, ethnic minority borrowers pay about .18% higher mortgage interest rates. However, when the lender is not among the top twenty five in terms of origination volume, our results indicate that smaller lenders charge .375% higher mortgage interest rates. These charges legally could be deemed disparate treatment under federal statutes, since they are unrelated to the life cycle characteristics of African American and Hispanic borrowers.

6 Conclusions

To investigate whether the use of machine-learning-based algorithmic underwriting has led to statistical discrimination among protected classes of borrowers, we analyze a mortgage data set that includes never-before-linked information at the loan-level on income, ethnicity, debt-to-income ratios, loan-to-value ratios, all contract terms, and indicators for whether the lender-of-record primarily used algorithmic underwriting. After controlling for observable differences, we find that African-American and Hispanic borrowers are almost 2% more likely to be rejected for a mortgage than other borrowers; conditional on obtaining a loan, they pay a slightly (0.18%) higher interest rate. These differences are less pronounced among lenders who utilize algorithmic underwriting and more pronounced for smaller lenders.

We make important methodological contributions to current legal debates concerning the "robust statistical" measurement of *disparate impact* as required by the courts and debates within the economic literature concerning the identification of illegitimate discriminatory use of protected characteristics in credit screening. Our identification strategy relies on the unique screening and grid pricing strategies of the GSES and allows us to test for disparate impact even in a setting in which borrowers select to apply for a mortgage and where not all life cycle variables, such as wealth, are observable. Since the standard of proof for a disparate impact claim is that a claimant must prove (i) that discrimination occurred and (ii) that no legitimate business necessity mandated the use of a sorting mechanism that discriminated, our empirical strategy successfully translates these requirements into economics in two parts. First, we successfully identified statistical discrimination in mortgage interest rates charged to borrowers via our application of the Oaxaca decomposition method. Second, we establish that the GSE underwriting and pricing structure insures that there is no omitted variable, unobservable, which could be part of a structural model that is a legitimate business necessity. We then establish, again through application of the Oaxaca decomposition, that lenders accept/reject decisions, where credit risk is a legitimate business necessity based on the life cycle characteristics of borrowers, also exhibits an important discriminatory component that is unexplained by legitimate life cycle characteristics. This component is comprised of differential treatment of ethnic minorities on average and on illegitimate differential weighting of the similar life cycle variables between African-American/Hispanic borrowers and Causian/Asian borrowers.

The implications of our findings for more algorithmic FinTech mortgage lending suggest that, in addition to the efficiency gains of these innovations, they may also serve to make the mortgage lending markets more accessible to African-American and Hispanic borrowers and provide these borrowers with fairer pricing. This positive evaluation of the net benefits of algorithmic underwriting bodes wells for the needed expansion of U.S. residential mortgage markets as they continue to recover from the 2009 crisis. On a more cautionary note, however, the discipline imposed by the GSE's underwriting and pricing requirements are likely to be an important reason for the observed small differentials in mortgages pricing between ethnic minorities and majority borrower populations. To date, this less well understood role of the GSEs has not been considered in GSE reform proposals, nor is it obvious how such a role could be supported within a fully privatized conventional conforming secondary mortgage market.

Figure 1: LLPA Adjustments from Sample Rate Sheet

Table 2: All Eligible Mortgages (Excluding MCM): LLPA by Credit Score/LTV

PRODUCT FEATURE	LLPAs by LTV Range								
	≤ 60.00%	60.01 – 70.00%	70.01 – 75.00%	75.01 – 80.00%	80.01 – 85.00%	85.01 – 90.00%	90.01 – 95.00%	95.01 – 97.00%	SFC
Representative Credit Score	Applicable for all mortgages with greater than 15 year terms For whole loans purchased on or before March 31, 2011, or loans delivered into MBS pools with issue dates of March 1, 2011 or earlier								
≥ 740	-0.250%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	N/A
720 – 739	-0.250%	0.000%	0.000%	0.250%	0.000%	0.000%	0.000%	0.000%	N/A
700 – 719	-0.250%	0.500%	0.500%	0.750%	0.500%	0.500%	0.500%	0.500%	N/A
680 – 699	0.000%	0.500%	1.000%	1.500%	1.000%	0.750%	0.750%	0.500%	N/A
660 – 679	0.000%	1.000%	2.000%	2.500%	2.250%	1.750%	1.750%	1.250%	N/A
640 – 659	0.500%	1.250%	2.500%	3.000%	2.750%	2.250%	2.250%	1.750%	N/A
620 – 639	0.500%	1.500%	3.000%	3.000%	3.000%	2.750%	2.750%	2.500%	N/A
< 620 (1)	0.500%	1.500%	3.000%	3.000%	3.000%	3.000%	3.000%	3.000%	N/A

Figure 2: GSE Framework Providing Comparable Risk Pricing to Lenders for within-Grid Mortgages

The figure represents two mortgage applications with the same maturity, the same credit score and the same loan-to-value (LTV) ratio. The interest rate a household sees on a mortgage consists of three parts. First, all mortgages within the same credit score-LTV face the same market price, which emerges from the Base Mortgage Rate coming from the primary market for loans to be securitized by the GSEs. Second, all mortgages within the same maturity-credit score-LTV face the same G-fee added onto the market price. The G-fee, or guarantee fee, is the fee that the GSEs assign to insure the mortgage against default. The third part of the pricing is discretionary to the originator and is determined by the lenders revenue model and profits from their proprietary credit scoring.

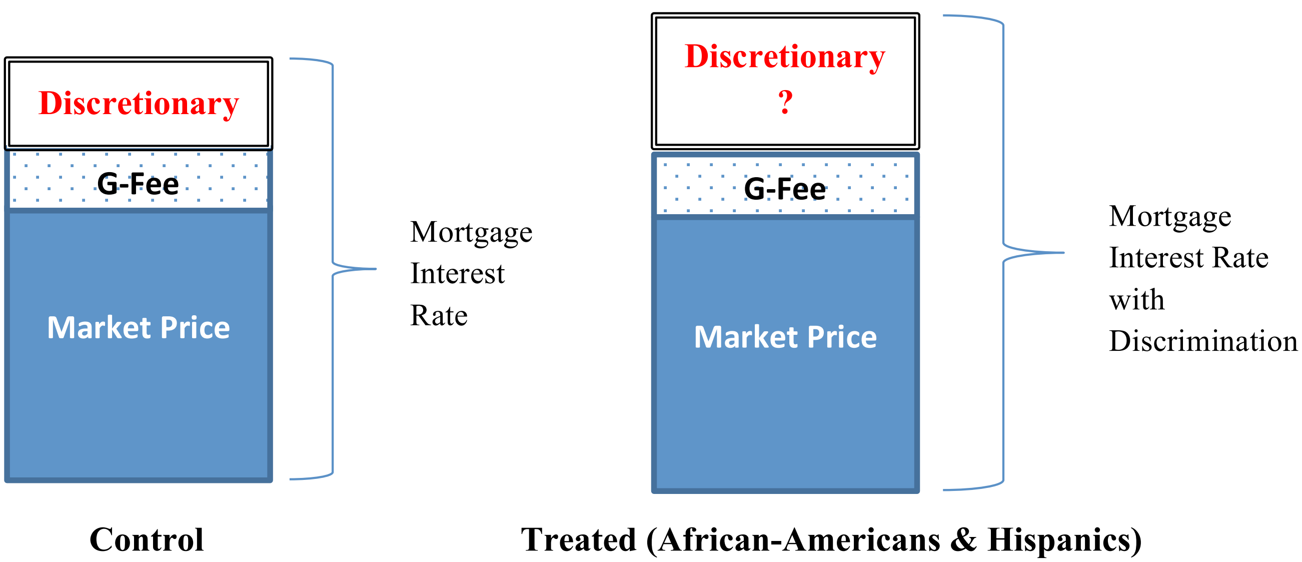
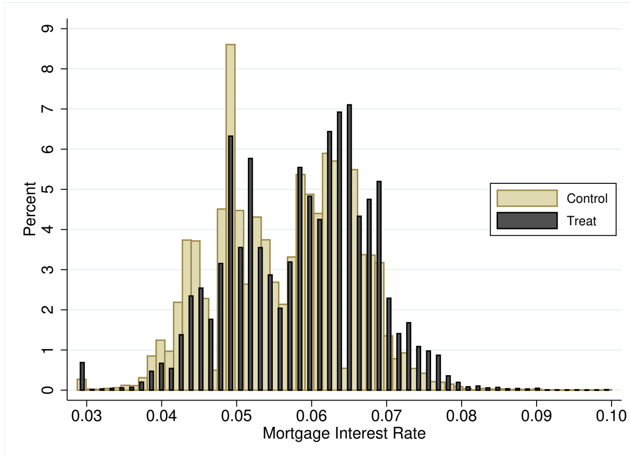
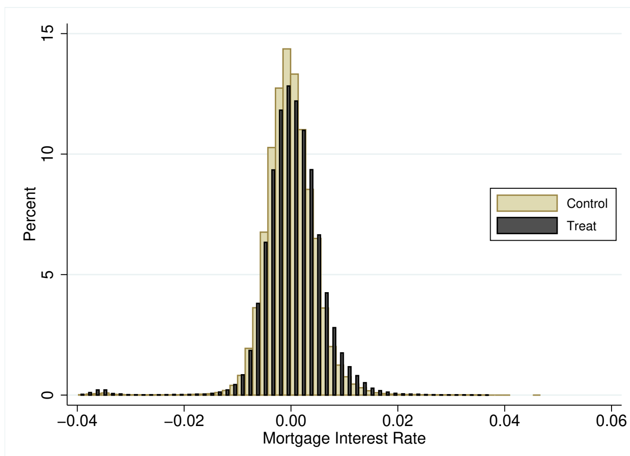


Figure 3: Borrower interest rates by treatment

(a) Raw data



(b) De-meaned by subtracting year-grid cell mean



(c) De-meaned by subtracting year-grid cell mean (debt sample)

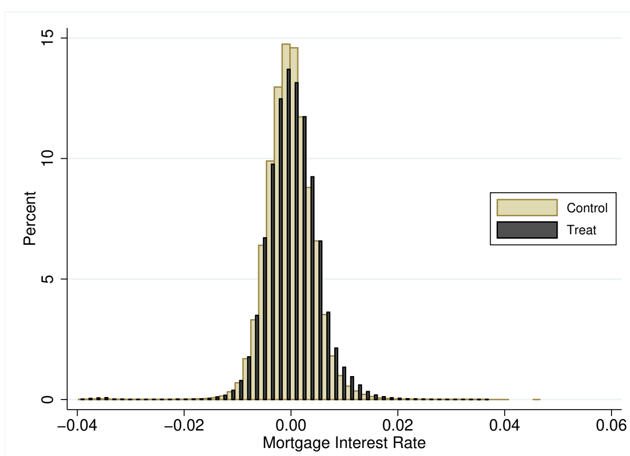


Figure 4: Comparison of loan application rejection probabilities for the raw and de-meaned to year buckets: for the control group (ethnicity $\neq$ African-American/Hispanic) versus the treatment group (ethnicity = African-American/Hispanic).

The figure compares the yearly loan application rejection probabilities (estimates for the raw and the demeaned to year buckets) for the full sample control group (ethnicity $\neq$ African-American/Hispanic) versus the treatment group (ethnicity = African-American/Hispanic).

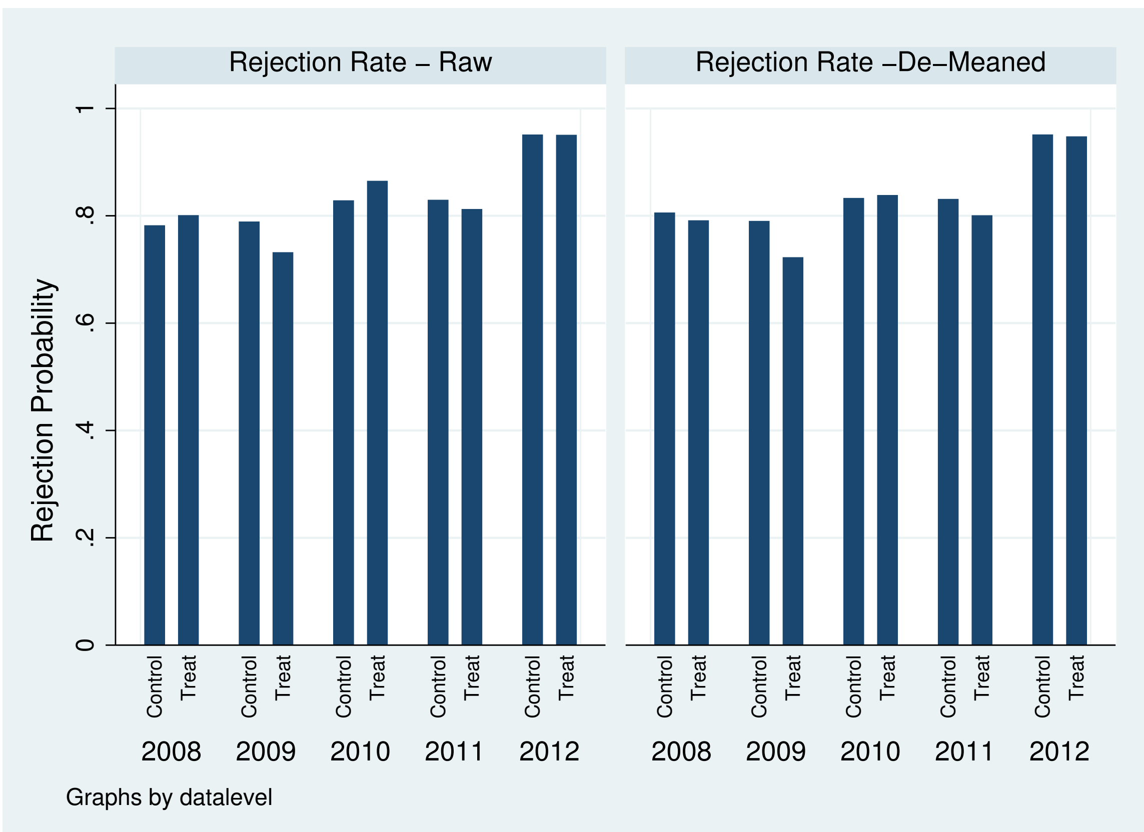


Figure 5: Discrimination in rejection rates for purchase mortgages by FICO and LTV

The figure show the discrimination coefficient from running Oaxaca (1973) regressions of mortgage rejection rates on LTV/FICO buckets.

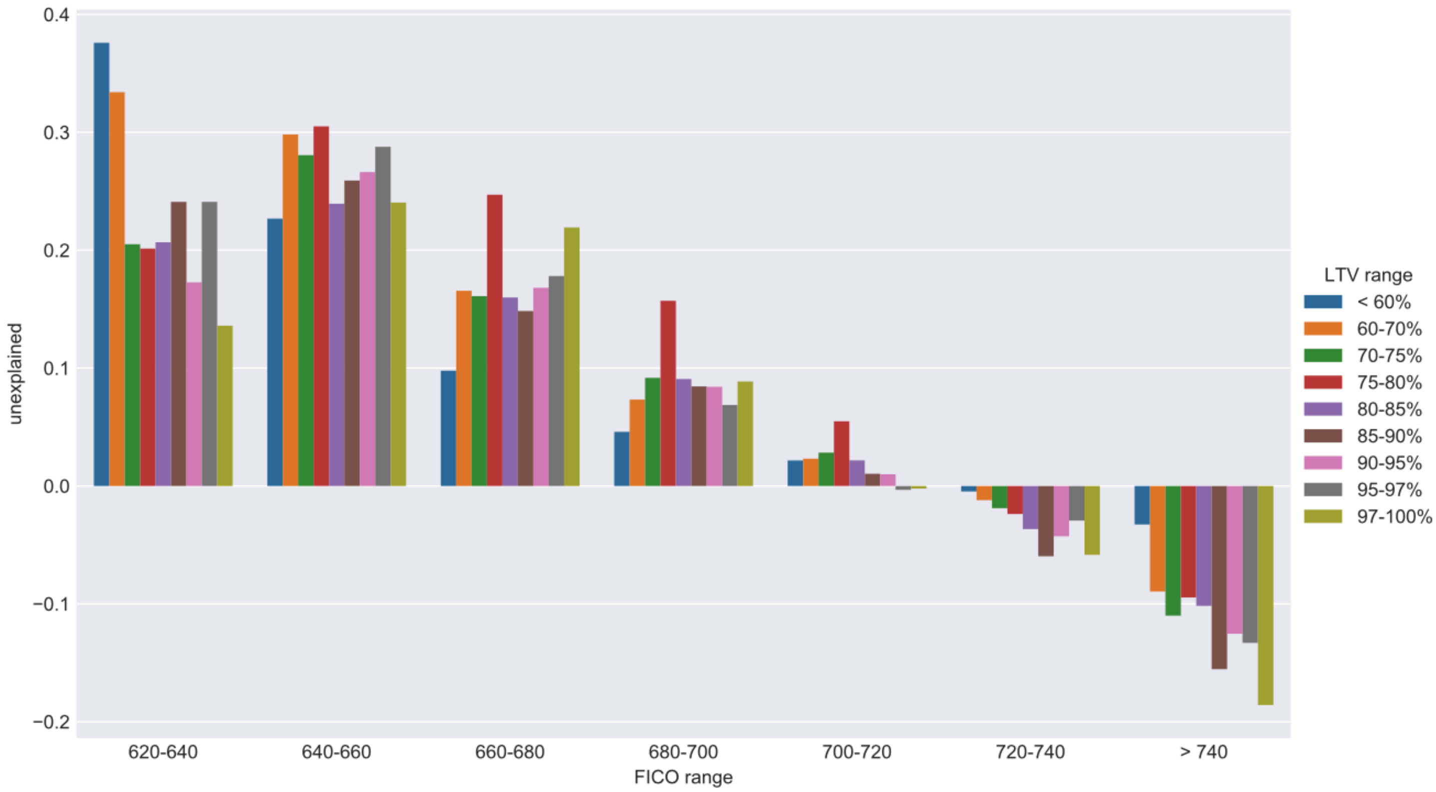


Figure 6: Discrimination in interest rates for purchase mortgage by FICO and LTV

The figures shows the discrimination coefficients from running Oaxaca (1973) regressions of mortgage interest rates on the life cycle variables for each of 72 LTV/FICO buckets. The estimated coefficients are then averaged by FICO buckets as shown in the figure.

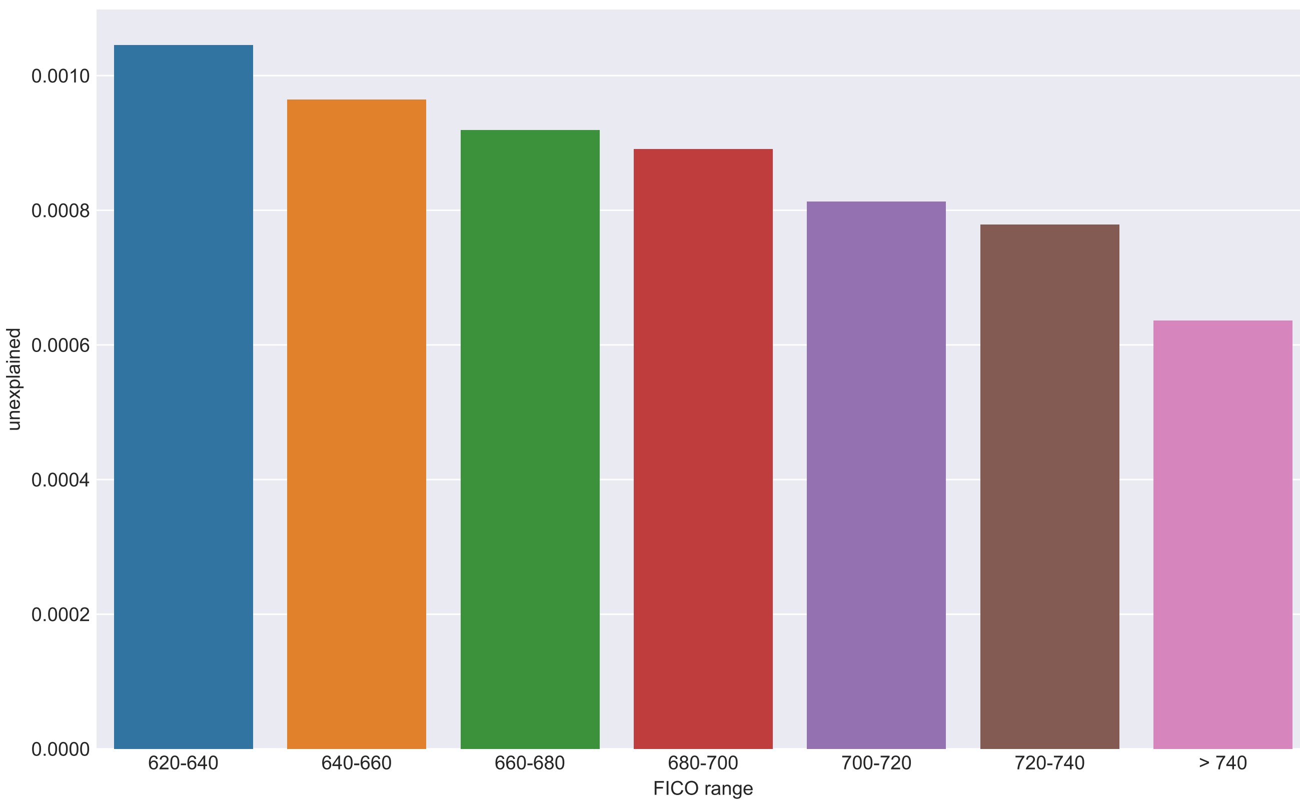


Table 1: Summary Statistics

Summary statistics for the originated and rejected loan samples. The information for the rejected loans was obtained from HMDA. As discussed in the Appendix, the credit scores for the rejected loans are proxied as the median vantage 3.0 score by census tract, as reported for all loans in the McDash data set. The loan-to-value ratios for rejected loan applications are proxied by the HMDA reported loan amount requested divided by the median house price in the loan census tract. The median house prices are measured as the assessed market value of all houses by census tract as reported in the Dataquick Assessor files by year. The originated loan data is from a merge of HMDA, Dataquick and McDash data sets, as discussed in the Appendix. The top 25 lenders are identified from annual loan origination data reported by *Inside Mortgage Finance*, for the years 2008 through 2012. The FinTech lenders were identified by lender name and were classified as FinTech due to their focus on on-line and algorithmic underwriting practices following the taxonomy introduced by Buchak et al. (2017).

Panel A: Rejected Applications (N = 5,253,561)					
	Mean	Standard Deviation	Minimum	Median	Maximum
Interest Rate %	--	--	--	--	--
Loan Amount $	158,874	81,730	30,000	142,000	429,000
Applicant Income $	80,791	66,418	10,000	64,000	999,000
Loan-to-Value	0.719	0.180	0.300	0.741	1.000
Credit Score	694	29	630.5	694.5	770
FinTech	0.025				
Top 25	0.638				
African_American/Hispanic	0.142				
Purchase=1; Refinance=0	0.355				

Panel B: Accepted Applications without Equifax match (N = 1,583,752)					
	Mean	Standard Deviation	Minimum	Median	Maximum
Interest Rate %	0.057	0.009	0.029	0.058	0.100
Loan Amount $	222,213	108,207	10,000	202,000	600,000
Applicant Income $	99,092	70,985	10,000	83,000	999,000
Loan-to-Value	0.748	0.147	0.300	0.781	1.000
Credit Score	719	36	631	725	770
FinTech	0.024				
Top 25	0.466				
African_American/Hispanic	0.159				
Purchase=1; Refinance=0	0.431				

Panel C: Accepted Applications with Equifax match (N = 1,121,395)					
	Mean	Standard Deviation	Minimum	Median	Maximum
Interest Rate %	0.056	0.009	0.029	0.055	10.0
Loan Amount $	226,266	108,121	10,000	207,000	600,000
Applicant Income $	98,887	69,291	10,000	83,000	999,000
Loan-to-Value	0.747	0.148	0.300	0.780	1.000
Credit Score	720	36	631	726	770
FinTech	0.025				
Top 25	0.475				
African_American/Hispanic	0.155				
Purchase=1; Refinance=0	0.412				
Debt (Other) Outstanding	42,490	59,495	0	23,840	3,630,719

Table 2: Rejection Discrimination: Main Results for Purchase Loans

Panel A: Discrimination Results within Grid

Column (1) shows OLS regression of the purchase mortgage rejection rate on the loan-to-value ratio, the credit score, and the log income of the borrower; column (2) includes month and bucket fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of purchase mortgage coupon rate on the same explanatory variables.

Panel B: Discrimination Results with Lender Fixed Effects

Column (1) shows OLS regression of the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, and the log income of the borrower; column (2) includes month, bucket fixed, and lender fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of purchase mortgage coupon rate on the same explanatory variables.

Panel A: Discrimination Results within Grid					
	Dependent Variable: Rejection				
	(1)	(2)	(3)	(4)	(5)
			Oaxaca Decomposition		
Model	OLS	OLS	Treatment	Explained	Unexplained
Discrimination	-0.0410***	-0.0307***	0.0199***		
	[0.000596]	[0.000551]	[0.000655]		
Loan-to-Value	-0.331***	-0.175***		-0.00239***	0.0238**
	[0.00121]	[0.00382]		[8.99e-05]	[0.00981]
Credit Score	-0.00452***	-0.00494***		0.0529***	0.614***
	[8.50e-06]	[3.63e-05]		[0.000481]	[0.0613]
Log Income	-0.0797***	-0.0486***		0.0106***	-0.0673***
	[0.000390]	[0.000370]		[9.75e-05]	[0.0103]
Observations	2,286,740	2,280,594	2,280,594		
R-squared	0.196	0.33			
Year FE	N	Y	Y		
Bucket FE	N	Y	Y		

Panel B: Discrimination Results With Lender Fixed Effects					
	Dependent Variable: Rejection				
	(1)	(2)	(3)	(4)	(5)
			Oaxaca Decomposition		
Model	OLS	OLS	Treatment	Explained	Unexplained
Discrimination	-0.0333***	-0.0220***	0.0179***		
	[0.000570]	[0.000542]	[0.000662]		
Loan-to-Value	-0.281***	-0.179***		-0.00231***	0.0194**
	[0.00115]	[0.00382]		[8.61e-05]	[0.00928]
Credit Score	-0.00396***	-0.00370***		0.0464***	0.519***
	[8.46e-06]	[3.51e-05]		[0.000443]	[0.0576]
Log Income	-0.0677***	-0.0565***		0.00940***	-0.0638***
	[0.000375]	[0.000361]		[9.19e-05]	[0.00984]
Observations	2,217,567	2,211,651	2,211,651		
R-squared	0.321	0.391			
Year FE	N	Y	Y		
Bucket FE	N	Y	Y		
Lender FE	Y	Y	Y		

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Rejection Discrimination: Results for Purchase Loans by Traditional Lenders vs. FinTech Lenders

Columns (1)–(3) show results of Oaxaca (1973) regression of the purchase mortgage rejection rate for the traditional (non-FinTech) lenders on the loan-to-value ratio, the credit score, the log income of the borrower and year and bucket fixed effects. Columns (4)–(6) show results of Oaxaca (1973) regression of purchase mortgage rejection rate for the FinTech lenders on the same explanatory variables.

	Dependent Variable: Rejection					
	(1)	(2)	(3)	(4)	(5)	(6)
	Traditional Lenders			FinTech Lenders		
	Oaxaca Decomposition			Oaxaca Decomposition		
Model	Treatment	Explained	Unexplained	Treatment	Explained	Unexplained
Discrimination	0.0208***			-0.0734***		
	[0.000656]			[0.00745]		
Loan-to-Value		-0.00237***	0.0239**		-0.00388***	-0.0351
		[8.99e-05]	[0.00983]		[0.00129]	[0.106]
Credit Score		0.0531***	0.613***		0.0379***	1.004
		[0.000484]	[0.0614]		[0.00385]	[0.643]
Log Income		0.0106***	-0.0601***		0.00401***	-0.296***
		[9.79e-05]	[0.0103]		[0.000887]	[0.110]
Observations	2,250,839			29,755		
Year FE	Y			Y		
Bucket FE	Y			Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Rejection Discrimination: Results for Purchase loans by Top 25 Lenders vs. Smaller Lenders

Columns (1)–(3) show results of Oaxaca (1973) regression of the purchase mortgage rejection rate for the top 25 lenders on the loan-to-value ratio, the credit score, the log income of the borrower and year and bucket fixed effects. Columns (4)–(6) show results of Oaxaca (1973) regression of purchase mortgage rejection rate for the smaller lenders on the same explanatory variables.

	Dependent Variable: Rejection					
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 25 Lenders			Small Lenders		
	Oaxaca Decomposition			Oaxaca Decomposition		
Model	Treatment	Explained	Unexplained	Treatment	Explained	Unexplained
Discrimination	0.00832***			0.0471***		
	[0.000746]			[0.00121]		
Loan-to-Value		-0.00161***	0.00447		-0.00424***	0.0605***
		[9.04e-05]	[0.0114]		[0.000206]	[0.0175]
Credit Score		0.0499***	0.560***		0.0554***	0.585***
		[0.000544]	[0.0709]		[0.000905]	[0.110]
Log Income		0.00929***	-0.0235*		0.0132***	-0.148***
		[0.000108]	[0.0120]		[0.000189]	[0.0181]
Observations	1,468,133			812,461		
Year FE	Y			Y		
Bucket FE	Y			Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Interest Rate Discrimination: Main Results for Purchase Loans

Panel A: Full Sample
Column (1) shows OLS regression of the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower, and the other consumer debt held by the borrower; column (2) includes month and bucket fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of purchase mortgage coupon rate on the same explanatory variables.
Panel B: Sample that Merges with Equifax Debt Data
Column (1) shows OLS regression of the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower, and the other consumer debt held by the borrower; column (2) includes month and bucket fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of purchase mortgage coupon rate on the same explanatory variables.

Panel A: Full Sample					
	Dependent Variable: Mortgage Interest Rate				
	(1)	(2)	(3)	(4)	(5)
			Oaxaca Decomposition		
Model	OLS	OLS	Treatment	Explained	Unexplained
Discrimination	0.00129***	0.000921***	0.00179***		
	[2.93e-05]	[1.85e-05]	[3.05e-05]		
Loan-to-Value	0.00488***	0.000681***		1.10e-05*	0.00134**
	[8.62e-05]	[0.000218]		[5.72e-06]	[0.000624]
Credit Score	-5.40e-05***	-1.04e-05***		9.15e-05***	0.00047
	[2.98e-07]	[8.27e-07]		[8.00e-06]	[0.00191]
Log Income	-0.000219***	-0.000346***		4.93e-05***	-0.00335***
	[1.83e-05]	[1.05e-05]		[1.89e-06]	[0.000358]
Observations	683,120	662,807	662,807		
R-squared	0.064	0.697			
Year FE	N	Y	Y		
Bucket FE	N	Y	Y		

Panel B: Sample that Merges with Equifax Debt Data					
	Dependent Variable: Mortgage Interest Rate				
	(1)	(2)	(3)	(4)	(5)
			Oaxaca Decomposition		
Model	OLS	OLS	Treatment	Explained	Unexplained
Discrimination	0.00130***	0.000839***	0.00171***		
	[3.49e-05]	[2.04e-05]	[3.62e-05]		
Loan-to-Value	0.00329***	0.000564**		7.44E-06	0.00148**
	[0.000103]	[0.000250]		[6.33e-06]	[0.000691]
Credit Score	-5.30e-05***	-1.02e-05***		8.91e-05***	0.00133
	[3.56e-07]	[9.23e-07]		[8.68e-06]	[0.00210]
Log Income	-0.000598***	-0.000533***		8.21e-05***	-0.00309***
	[2.25e-05]	[1.23e-05]		[2.43e-06]	[0.000412]
Log Debt	0.000457***	7.33e-05***		-1.62e-05***	-0.000126
	[6.61e-06]	[3.62e-06]		[1.03e-06]	[9.41e-05]
Observations	462,126	450,832	450,832		
R-squared	0.072	0.733			
Debt Covariates	Y	Y	Y		
Year FE	N	Y	Y		
Bucket FE	N	Y	Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Interest Rate Discrimination: Results for Purchase Loans by Traditional Lenders vs. FinTech Lenders

Columns (1)–(3) show the results of Oaxaca (1973) regression of the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower, and month and bucket fixed effects for the traditional (non-FinTech) lenders. Columns (4)–(6) run the same regression for the FinTech lenders.

	Dependent Variable: Mortgage Interest Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
	Traditional Lenders			FinTech Lenders		
	Oaxaca Decomposition			Oaxaca Decomposition		
Model	Treatment	Explained	Unexplained	Treatment	Explained	Unexplained
Discrimination	0.00174***			0.00037		
	[3.64e-05]			[0.000268]		
Loan-to-Value		7.03E-06	0.00155**		0.000111	-0.00918
		[6.32e-06]	[0.000694]		[7.83e-05]	[0.00674]
Credit Score		8.92e-05***	0.00109		6.87E-05	0.0102
		[8.75e-06]	[0.00212]		[6.57e-05]	[0.0162]
Log Income		8.08e-05***	-0.00299***		0.000157***	-0.0107***
		[2.44e-06]	[0.000416]		[2.50e-05]	[0.00316]
Log Debt		-1.61e-05***	-0.000149		-1.67e-05*	0.00125*
		[1.04e-06]	[9.49e-05]		[9.35e-06]	[0.000724]
Observations	444,204			6,628		
Year FE	Y			Y		
Bucket FE	Y			Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Interest Rate Discrimination: Results for Purchase Loans by Top 25 Lenders vs. Smaller Lenders

Columns (1)–(3) show the results of Oaxaca (1973) regressions the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower and month and bucket fixed effects for the large (top 25) lenders, while Columns (4)–(6) run the same regressions for the smaller (non-top-25) lenders.

	Dependent Variable: Mortgage Interest Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 25 Lenders			Small Lenders		
	Oaxaca Decomposition			Oaxaca Decomposition		
Model	Treatment	Explained	Unexplained	Treatment	Explained	Unexplained
Discrimination	0.00183***			0.00159***		
	[5.05e-05]			[5.16e-05]		
Loan-to-Value		-1.96E-06	0.00242**		1.67e-05*	0.000582
		[8.62e-06]	[0.000959]		[9.28e-06]	[0.000993]
Credit Score		9.15e-05***	0.00439		8.57e-05***	-0.00127
		[1.25e-05]	[0.00301]		[1.21e-05]	[0.00293]
Log Income		8.04e-05***	-0.00338***		7.99e-05***	-0.00276***
		[3.50e-06]	[0.000599]		[3.35e-06]	[0.000571]
Log Debt		-1.38e-05***	4.85E-06		-1.81e-05***	-0.000251*
		[1.40e-06]	[0.000136]		[1.50e-06]	[0.000130]
Observations	211,605			239,227		
Year FE	Y			Y		
Bucket FE	Y			Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

References

Abowd, J., F. Kramarz, and D. Margolis, 1999, High wage workers and high wage firms, *Econometrica* 67, 251–333.

Aigner, D. J., and G. Cain, 1977, Statistical theories of discrimination in the labor market, *Industry and Labor Relations Review* 30, 175–187.

Arrow, K., 1973, Higher education as a filter, *Journal of Public Economics* 2, 193–216.

Black, H., R. L. Schweitzer, and L. Mandell, 1978, Wage discrimination: Reduced form and structural estimates, *American Economic Review, Papers and Proceedings* 68, 186–191.

Black, H. A., and R. Schweitzer, 1977, A canonical analysis of mortgage lending terms: Testing for lending discrimination at a commercial bank, *Urban Studies* 22, 13–19.

Blinder, A. S., 1973, Wage discrimination: Reduced form and structural estimates, *Journal of Human Resources* 8, 436–455.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru, 2017, Fintech, regulatory arbitrage, and the rise of shadow banks, Working paper, University of Chicago.

Card, D., A. R. Cardoso, and P. Kline, 2016, Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women, *Quarterly Journal of Economics* 131, 633–686.

Elenev, V., T. Landvoigt, and S. van Nieuwerburgh, 2016, Phasing out the GSEs, *Journal of Monetary Economics* 81, 11–132.

Engel, K. C., and P. A. McCoy, 2011, *The Subprime Virus: Reckless Credit, Regulatory Failure, and Next Steps* (Oxford University Press, New York).

Federal Housing Finance Administration, 2009, Fannie Mae and Freddie Mac single-family guarantee fees in 2007 and 2008.

Federal Housing Finance Administration, 2010, Fannie Mae and Freddie Mac single-family guarantee fees in 2008 and 2009.

Federal Housing Finance Administration, 2011, Fannie Mae and Freddie Mac single-family guarantee fees in 2009 and 2010.

Federal Housing Finance Administration, 2012, Fannie Mae and Freddie Mac single-family guarantee fees in 2010 and 2011.

Federal Housing Finance Administration, 2013, Fannie Mae and Freddie Mac single-family guarantee fees in 2012.

Fortin, N., T. Lemieux, and S. Firpo, 2011, Decomposition methods in economics, in O. Ashenfelter, and D. Card, eds., *Handbook of Labor Economics*, volume 4A, chapter 1 (Elsevier, Amsterdam).

Fuster, A., L. Goodman, D. Lucca, L. Madar, L. Molloy, and P. Willen, 2013, The rising gap between primary and secondary mortgage rates, *Federal Reserve Bank of New York Economic Policy Review* 19.

Fuster, A., and P. Willen, 2010, $1.25 trillion is still real money: Some facts about the effects of the Federal Reserve's mortgage market investments, Public Policy Discussion Paper 10-4, Federal Reserve Bank of Boston.

Gano, A., 2017, Disparate impact and mortgage lending: A beginner's guide, *University of Colorado Law Review* 88, 1109–1166.

Ghent, A. C., R. Hernández-Murillo, and M. T. Owyang, 2014, Differences in subprime loan pricing across races and neighborhoods, *Regional Science and Urban Economics* 48, 199–215.

Glassman, A. M., and S. Verna, 2016, Disparate impact one year after inclusive communities, *Journal of Affordable Housing & Community Development Law* 25, 12–24.

Hastie, T., R. Tibshirani, and J. Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition (Springer, New York).

Haughwout, A., C. Mayer, and J. Tracy, 2009, Subprime mortgage pricing: The impact of race, ethnicity, and gender on the cost of borrowing, Staff Report 368, Federal Reserve Bank of New York.

James, G., D. Witten, T. Hastie, and R. Tibshirani, 2015, *An Introduction to Statistical Learning with Applications in R* (Springer, New York).

Kaye, D., 1982, Statistical evidence of discrimination, *Journal of the American Statistical Association* 77, 773–783.

Kerr, W. R., 2008, Ethnic scientific communities and international technology diffusion, *Review of Economics and Statistics* 90, 518–537.

Kerr, W. R., and W. F. Lincoln, 2010, The supply side of innovation: H-1B visa reforms and U.S. ethnic invention, *Journal of Labor Economics* 28, 473–508.

Kline, P., 2011, Oaxaca-Blinder as a reweighting estimator, *American Economic Review: Papers & Proceedings* 101, 532–537.

Maddala, G., and R. P. Trost, 1982, On measuring discrimination in loan markets, *Housing Finance Review* 1, 693–709.

Oaxaca, R., 1973, Male-female wage differentials in urban labor markets, *International Economic Review* 14, 693–709.

Oaxaca, R. L., and M. R. Ransom, 1999, Identification in detailed wage decompositions, *Review of Economics and Statistics* 81, 154–157.

Phelps, E. S., 1972, Money, public expenditure and labor supply, *Journal of Economic Theory* 5, 69–78.

Rachlis, M., and A. Yezer, 1993, Serious flaws in statistical tests for discrimination in mortgage markets, *Journal of Housing Research* 4, 315–336.

Sandler, A. L., and J. Biran, 1995, The improper use of statistics in mortgage lending discrimination actions, in A. Yezer, ed., *Fair Lending Analysis* (American Bankers Association, Washington, DC).

Shafer, R., and H. Ladd, 1981, *Discrimination in Mortgage Lending* (MIT Press, Cambridge, MA).

Słoczyński, T., 2015, The Oaxaca-Blinder unexplained component as a treatment effects estimator, *Oxford Bulletin of Economics and Statistics* 77, 588–604.

Vickery, J. I., and J. Wright, 2013, TBA trading and liquidity in the agency MBS market, *Economic Policy Review* 19.

A Algorithm for Merging Mortgage Data Sets

Since there are no unique mortgage loan identifiers in the U.S., we develop an algorithm using machine learning techniques to match loans found in two independent datasets: the McDash dataset, which contains loan-level data compiled by Black Night Financial Services, and the CoreLogic DataQuick dataset, which provides detailed property transaction and ownership information in addition to a time series history of all recorded mortgage lien events such new mortgage originations, prepayments, REO, foreclosure, short sales, and arms-length sales and loan payoffs. The algorithm developed to match these two data sets relies on matching distinct loan characteristics, e.g. origination date, loan amount, and termination and distress events so as to precisely match each loan with its counterpart in the other dataset using a modified k-nearest neighbor classifier (see Hastie, Tibshirani, and Friedman, 2009; James, Witten, Hastie, and Tibshirani, 2015).

A.1 Merge process for newly originated fixed rate GSE loans

There are two steps to our merging process. The first step is to serialize the Dataquick data into a record format in which each property is assigned a full event history string for each mortgage lien and the priority and performance of these mortgage loan positions. Dataquick provides very comprehensive geographic coverage for mortgage originations and terminations at the property-level for most five digit zip codes and most census tracks in the U.S. It accounts for about 90 percent of property Assessor's Pin Numbers and all the mortgage and lien recording for each property all sourced from public records. The second stage of the merging process is to employ functions for k-nearest neighbor algorithms using *sklearn.neighbors* in Python to fit radial kernels using *BallTree*. The k-neighbors classifier implements learning based on the 25 nearest neighbors in the corresponding zip code within the McDash mortgage data that also records the loan contract features and a loan-level string of performance characteristics. We represent each loan in each data set with a thirteen element vector that includes: 1) the original loan balance; 2) the lien position, 3) the origination date of the loan, 4) the ending date of the loan, 5) the foreclosure date of the loan (maybe null), the prepayment date of the loan (maybe null), 6) the appraised market value of the property, 7) the loan purpose (refinance or purchase), 8) loan distress dates (may be null), 9) loan REO date (may be null), 10) loan liquidation date (may be null), 11) short sale indicator variable (may be null), 12) interest rate type (fixed or variable loans), 13) property transaction value if there is a sale. Each of these elements are assigned a *category subscore* that is normalized to a float value between 0 and 1. Each sub-score is them squared to achieve a greater penalty for matches on key elements such as the loan amount. The category subscore is then scaled by a *category factor* which represents the categories importance to the match quality relative to other elements uses in the match. Each category factor is an integer between 0 and100 and the sum of the category factors is equal to 100. Our scoring algorithm (*get.score* in Python's sklearn) takes into account the 13 different elements of each matched pair of loans in order to calculate a score. The score roughly corresponds to the estimated error for each match, measured in hundredths of a percent. Thus, a match score of 1689 corresponds to a 16.89% chance of an incorrect match, or an 83.11% confidence in the match. We use only good quality matches with scores of 2000 or less. The merge rates for the Dataquick to McDash data sets using the modified $k - nearestneighbor$ algorithm for loans originated between 2008 and 2012 is 14.76 million loans. of these 8.8 million loans are GSE loans 90% of those are good merges.

Our prior machine learning strategy is less applicable for the merge of the HMDA data to McDash data, because we have only origination data in HMDA and as well as a greatly reduced set of loan characteristics at origination including: 1) the regulator type, 2) the loan type (conventional), 3) property type (1 to 4 single family residential properties, 4) loan purpose (refinance or purchase), 5) occupancy status, 6) original loan amount, 7) MSA, state, county and census tract, 8) self reported borrower and co-borrower ethnicity (African-American, Hispanic, Asian, Caucasian, unknown), 9) borrower/coborrower gender, 10) borrower annual income, 11) year of origination, 12) denial reason if loan application is rejected, and 13) lender name. For this merge, we instead we standardize the lenders names between Dataquick and HMDA and then merge these data sets using lender names, loan amount, lien type, and the loan purpose fields. Of the 30.6 million originations in the HMDA data sets and about 10.4 million GSE loans, we successfully merged 60% of these loans. We then merged the Dataquick to McDash merged data to the Dataquick to HMDA merged data using the crosswalk developed with the k-nearest neighbor algorithm and we obtained a final data set of 3.47 million loans that are single family fixed rate GSE loans originated between 2008 and 2012.

A.2 The Equifax-enhanced subsample of originations

To obtain a final data that includes the full spectrum of underwriting characteristics that would have been available to the lender, we again merge the HMDA/Dataquick/McDash data set of fixed rate GSE loans that were originated between 2008 and 2012 to a random sample of McDash loans that are merged to Equifax data. The Equifax-enhanced originated loan sample includes other consumer credit positions of the borrowers such as: the total sum of retail, consumer finance and bank card balances; total student loan debt, total auto loan debt (sum or auto finance and auto bank debt); age of the borrower, and Vantage 3.0 score.

A.3 The HMDA sample of rejected conventional loans

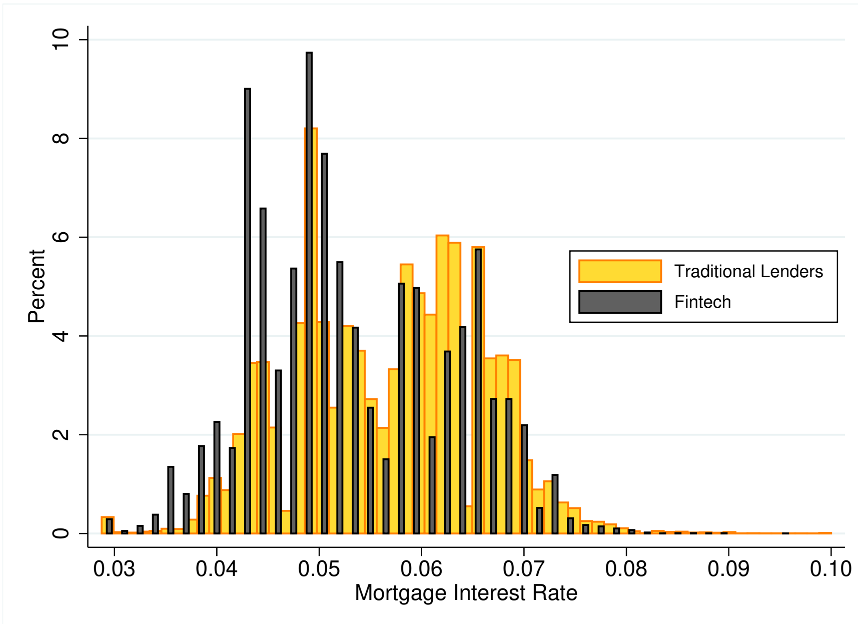
The second important class of loans in our data set includes all of the conventional conforming loans in HMDA for 1-4 family residential borrowers whose loan applications were either denied by the originator, approved but not accepted by the borrower, withdrawn by the applicant, or the loan application file was closed for incompleteness. These data include information on: 1) the regulator type, 2) the loan type (conventional), 3) property type (1 to 4 single family residential properties, 4) loan purpose (refinance or purchase), 5) occupancy status, 6) original loan amount, 7) MSA, state, county and census tract, 8) self reported borrower and co-borrower ethnicity (African-American, Hispanic, Asian, Caucasian, unknown), 9) borrower/co-borrower gender, 10) borrower annual income, 11) year of origination, 12) denial reason if loan application is rejected, and 13) lender name. They also include information on demographic and minority representation in the census tract in which the collateral on the loan is located. These variables include: the Federal Financial Institutions Examination Council (FFIEC) tract median family income to MSA median family income as a percentage, median family income for tract in thousands of dollars, tract population in thousands, tract minority population as a percentage, tract number of owner occupied units in thousands,tract number of 1- to 4-Family units in thousands.

HMDA does not include information on the credit score or the loan-to-value ratio of the rejected loan application files. For this reason, we proxy for the loan-to-value ratio by computing the mean assessed market value of all houses in each census tract in the U.S. using the panel of assessor's data from Dataquick. For each assessment we also have the year of the assessment. We then compute the ratio of the requested loan balance to the median value of all the homes in the appropriate census tract and year as reported by Dataquick to compute an estimated loan-to-value ratio for the rejected loan application. We proxy for the applicant's unobserved credit scores by using the McDash Vantage 3.0 score to compute the median credit score for each census track reported in McDash. The median census tract credit score is applied as a proxy for the credit score of the rejected loan.

B Robustness: Loan distribution by lender type

Figure 7: Borrower interest rates by treatment, FinTech vs. traditional lenders

(a) Raw data



(b) De-meaned by subtracting year-grid cell mean

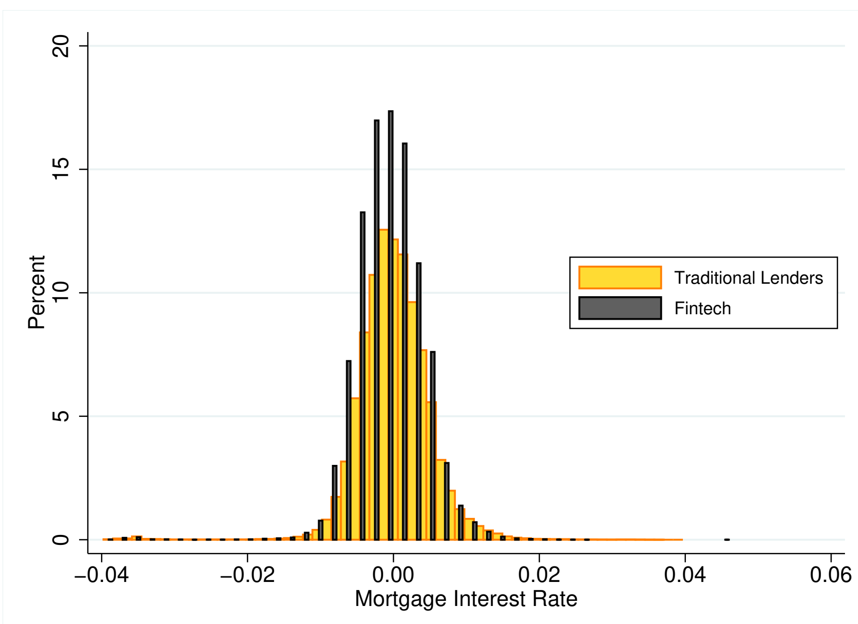
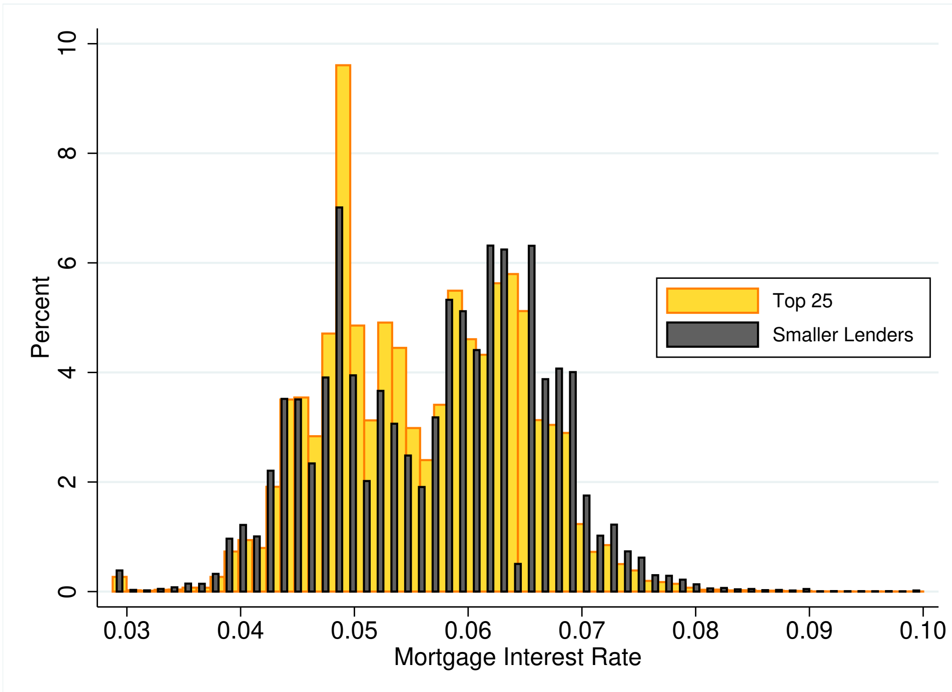
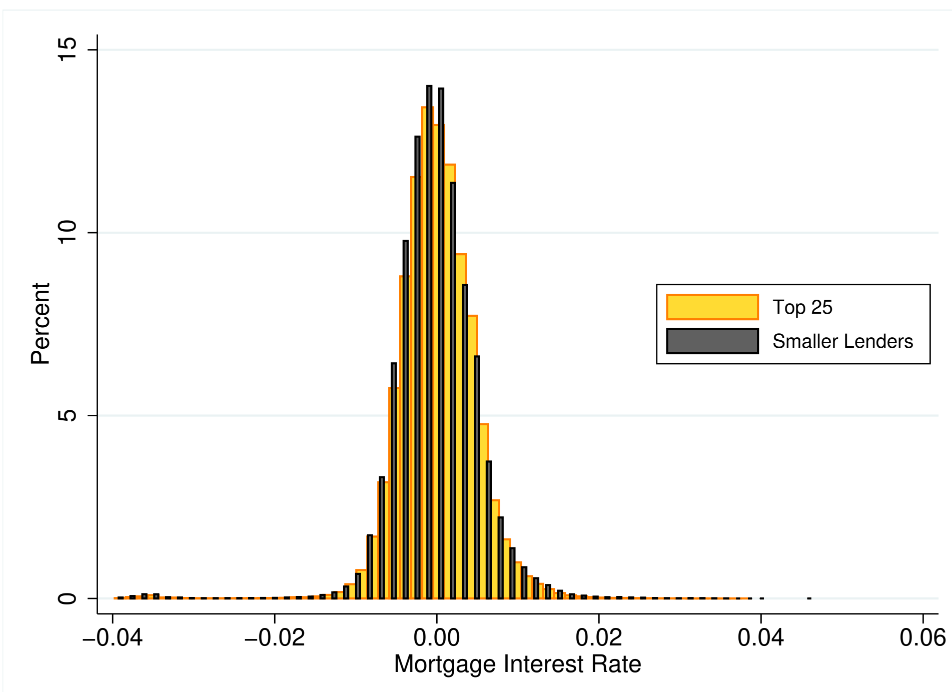


Figure 8: Borrower interest rates by treatment, large vs. small lenders

(a) Raw data



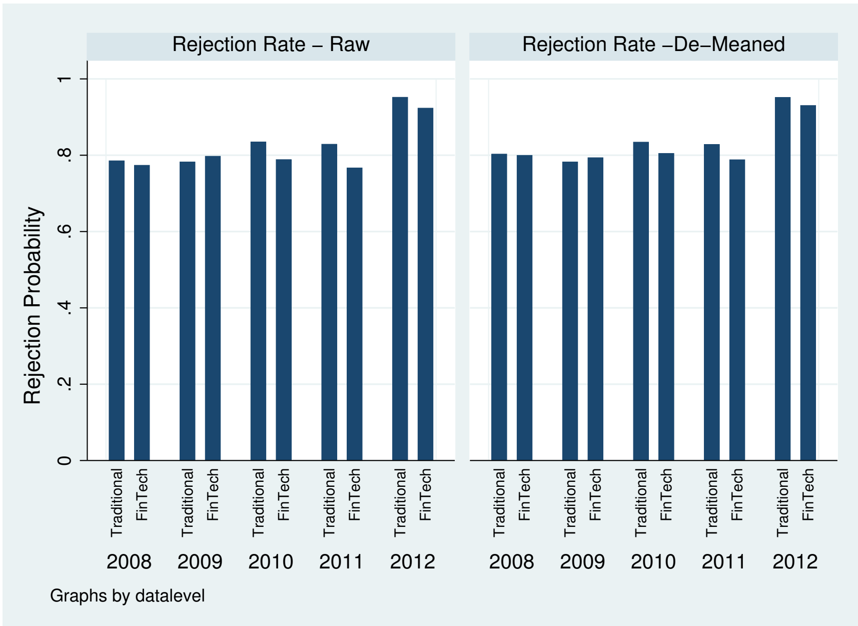
(b) De-meaned by subtracting year-grid cell mean



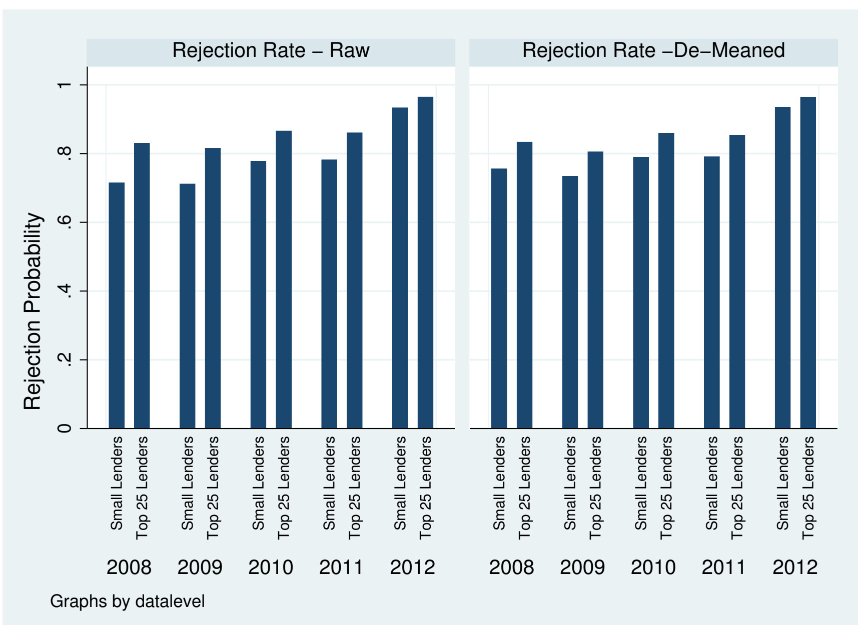
C Robustness: Rejection rates by lender type

Figure 9: Rejection rates by lender type

(a) FinTech versus traditional



(b) Large versus small



D Robustness: Results for Refinance Loans

Table 8: Interest Rate Discrimination: Main Results for Refi Mortgages

Panel A: Full Sample
Column (1) shows OLS regression of the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower, and the other consumer debt held by the borrower; column (2) includes month and bucket fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of the refinance mortgage coupon rate on the same explanatory variables.
Panel B: Sample that Merges with Equifax Debt Data
Column (1) shows OLS regression of the purchase mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower, and the other consumer debt held by the borrower; column (2) includes month and bucket fixed effects. Columns (3)–(5) show results of Oaxaca (1973) regression of the refinance mortgage coupon rate on the same explanatory variables.

Panel A: Full Sample					
	Dependent Variable: Mortgage Interest Rate				
	(1)	(2)	(3)	(4)	(5)
			Oaxaca Decomposition		
Model	OLS	OLS	Treatment	Explained	Unexplained
Discrimination	0.00204***	0.000212***	0.00296***		
	[2.58e-05]	[1.65e-05]	[2.70e-05]		
Loan-to-Value	0.00423***	-0.000164		4.98E-07	1.25E-05
	[6.08e-05]	[0.000109]		[4.08e-07]	[0.000270]
Credit Score	-7.14e-05***	-1.63e-05***		0.000151***	0.00201
	[2.54e-07]	[7.01e-07]		[7.19e-06]	[0.00172]
Log Income	-0.00185***	-0.000641***		9.86e-05***	-0.00135***
	[1.60e-05]	[9.22e-06]		[1.82e-06]	[0.000337]
Observations	900,632	896,164	896,164		
R-squared	0.125	0.705			
Year FE	N	Y	Y		
Bucket FE	N	Y	Y		

Panel B: Sample that Merges with Equifax Debt Data					
	Dependent Variable: Mortgage Interest Rate				
	(1)	(2)	(3)	(4)	(5)
			Oaxaca Decomposition		
Model	OLS	OLS	Treatment	Explained	Unexplained
Discrimination	0.00208***	0.000308***	0.00297***		
	[2.90e-05]	[1.67e-05]	[3.06e-05]		
Loan-to-Value	0.00378***	4.79E-05		8.70E-08	0.000297
	[6.79e-05]	[0.000120]		[7.68e-07]	[0.000267]
Credit Score	-7.12e-05***	-1.65e-05***		0.000143***	0.000133
	[2.87e-07]	[7.44e-07]		[7.42e-06]	[0.00173]
Log Income	-0.00210***	-0.000720***		0.000115***	-0.000511
	[1.81e-05]	[9.95e-06]		[2.16e-06]	[0.000345]
Log Debt	0.000325***	0.000119***		-1.17e-05***	-6.09E-05
	[5.71e-06]	[3.21e-06]		[8.30e-07]	[8.98e-05]
Observations	659,269	656,459	656,459		
R-squared	0.138	0.742			
Debt Covariates	Y	Y	Y		
Year FE	N	Y	Y		
Bucket FE	N	Y	Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Interest Rate Discrimination: Results for Refinance Mortgages by Traditional Lenders vs. FinTech Lenders

Columns (1)–(3) show the results of Oaxaca (1973) regression of the refinance mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower, and month and bucket fixed effects for the traditional (non-FinTech) lenders. Columns (4)–(6) run the same regression for the FinTech lenders.

	Dependent Variable: Mortgage Interest Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
	Traditional Lenders			FinTech Lenders		
	Oaxaca Decomposition			Oaxaca Decomposition		
Model	Treatment	Explained	Unexplained	Treatment	Explained	Unexplained
Discrimination	0.00299***			0.00124***		
	[3.09e-05]			[0.000196]		
Loan-to-Value		1.96E-07	0.000288		3.27E-06	0.00114
		[7.87e-07]	[0.000270]		[4.24e-06]	[0.00181]
Credit Score		0.000145***	0.00048		8.18e-05***	-0.0108
		[7.60e-06]	[0.00175]		[2.80e-05]	[0.0100]
Log Income		0.000115***	-0.000437		7.06e-05***	-0.00351*
		[2.21e-06]	[0.000350]		[1.12e-05]	[0.00211]
Log Debt		-1.18e-05***	-6.61E-05		-3.67E-06	0.00043
		[8.38e-07]	[9.10e-05]		[6.08e-06]	[0.000554]
Observations	635,516			20,943		
Year FE	Y			Y		
Bucket FE	Y			Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Interest Rate Discrimination: Results for Refinance Mortgages by Top 25 Lenders vs. Smaller Lenders

Columns (1)–(3) show the results of Oaxaca (1973) regressions the refinance mortgage coupon rate on the loan-to-value ratio, the credit score, the log income of the borrower and month and bucket fixed effects for the large (top 25) lenders, while Columns (4)–(6) run the same regressions for the smaller (non-top-25) lenders.

	Dependent Variable: Mortgage Interest Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 25 Lenders			Small Lenders		
	Oaxaca Decomposition			Oaxaca Decomposition		
Model	Treatment	Explained	Unexplained	Treatment	Explained	Unexplained
Discrimination	0.00227***			0.00375***		
	[4.07e-05]			[4.47e-05]		
Loan-to-Value		4.48E-08	0.000119		-7.95E-07	0.000438
		[5.51e-07]	[0.000344]		[1.58e-06]	[0.000410]
Credit Score		0.000119***	-0.00198		0.000171***	0.00219
		[9.12e-06]	[0.00225]		[1.20e-05]	[0.00261]
Log Income		0.000114***	-0.00124***		0.000113***	0.000708
		[3.02e-06]	[0.000441]		[3.04e-06]	[0.000534]
Log Debt		-1.15e-05***	0.000212*		-1.11e-05***	-0.000305**
		[1.29e-06]	[0.000120]		[1.03e-06]	[0.000133]
Observations	315,935			340,524		
Year FE	Y			Y		
Bucket FE	Y			Y		

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1